



UvA-DARE (Digital Academic Repository)

Disciplinary writing

Four empirical studies on historical and philosophical literacy

Holdinga, C.C.

Publication date

2023

Document Version

Final published version

[Link to publication](#)

Citation for published version (APA):

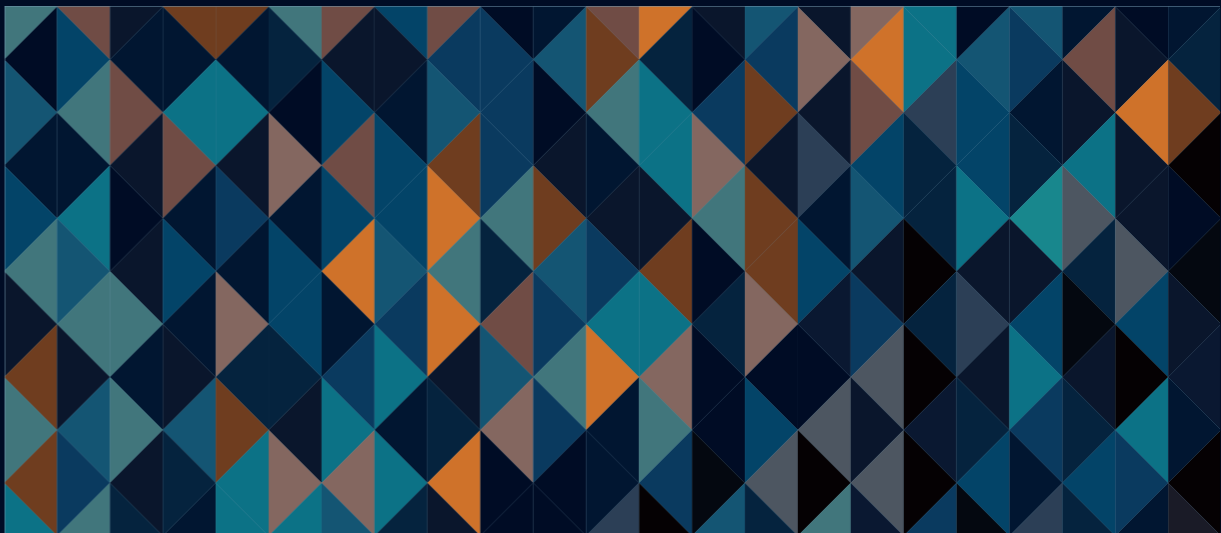
Holdinga, C. C. (2023). *Disciplinary writing: Four empirical studies on historical and philosophical literacy*. [Thesis, externally prepared, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

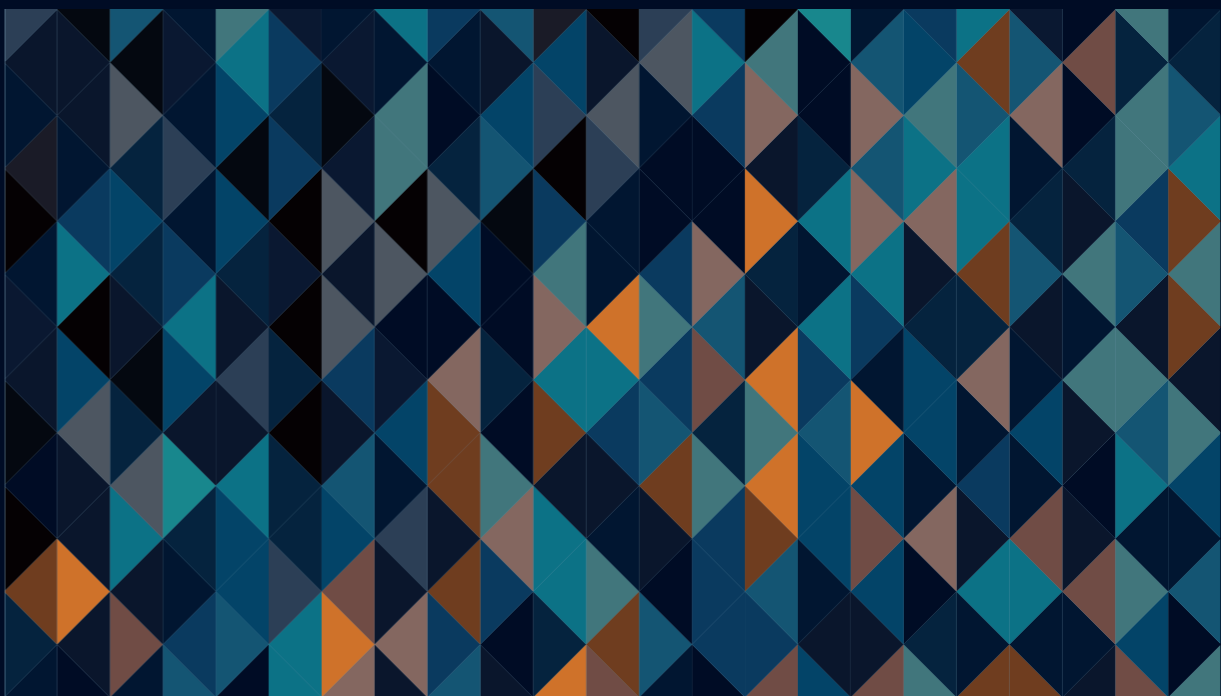
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.



DISCIPLINARY WRITING

Four empirical studies
on historical and
philosophical literacy

LIEKE HOLDINGA



DISCIPLINARY WRITING

LIEKE HOLDINGA

DISCIPLINARY WRITING
FOUR EMPIRICAL STUDIES ON HISTORICAL AND
PHILOSOPHICAL LITERACY



UNIVERSITY OF AMSTERDAM
Research Institute of Child Development
and Education

Disciplinary Writing
Four Empirical Studies on Historical and Philosophical Literacy
ISBN: 978-94-6483-459-8

Cover design: Patrick Kint
Printing: Ridderprint

Copyright © 2023 Lieke Holdinga
All rights reserved. No part of this thesis may be reproduced, stored, or transmitted in any way or by any means without the prior permission of the author, or when applicable, of the publishers of the scientific papers.

The research project presented in this publication was supported by the Dudo-Alpha Sustainable Humanities Programme, grant number DA2-2018-14.



DISCIPLINARY WRITING
FOUR EMPIRICAL STUDIES ON HISTORICAL AND
PHILOSOPHICAL LITERACY

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 17 november 2023, te 14.00 uur

door

Cornelia Caroline Holdinga
geboren te De Bilt

Promotiecommissie

Promotor: prof. dr. G. C. W. Rijlaarsdam
Universiteit van Amsterdam

Co-promotor: dr. J. P. van Drie
Universiteit van Amsterdam

Overige leden: prof. dr. M. L. L. Volman
Universiteit van Amsterdam

prof. dr. C. A. M. van Boxtel
Universiteit van Amsterdam

prof. dr. F. Kuiken
Universiteit van Amsterdam

prof. dr. S. De La Paz
University of Maryland

dr. D. W. Galbraith
University of Southampton

Faculteit der Maatschappij- en Gedragwetenschappen

TABLE OF CONTENTS

PROLOOG	Schrijvend leren in de vakles: effecten van schrijfprocesinstructie	7
CHAPTER 1	General Introduction	21
CHAPTER 2	Students' Disciplinary Writing Processes: A Think-Aloud Study	31
CHAPTER 3	Writing to Learn History: An Instructional Design Study	65
CHAPTER 4	Implementing Writing Process Instruction in the History Classroom: A Quasi-Experimental Study	99
CHAPTER 5	Fostering Philosophy Teachers' Disciplinary Writing Practice: A Multiple-Case Design Study	131
CHAPTER 6	General Discussion	169
REFERENCES		199
AUTHOR INDEX		211
APPENDICES		217
SUMMARY		237
SAMENVATTING		247
CONTRIBUTIONS OF AUTHORS		257
PUBLICATIONS AND PRESENTATIONS		259
ABOUT THE AUTHOR		261
DANKWOORD		263
DISSERTATION SERIES		265

PROLOOG

SCHRIJVEND LEREN IN DE VAKLES: EFFECTEN VAN SCHRIJFPROCESINSTRUCTIE

Dit proefschrift schreef ik in het Engels, om bij te kunnen dragen aan de internationale wetenschappelijke kennis rondom het ontwikkelen van schrijfvaardigheid in het onderwijs. Een belangrijk doel van het uitvoeren van dit onderzoek was echter om het onderwijs te verbeteren, al was het maar een fractie. Omwille van dat laatste doel schreef ik deze proloog, speciaal voor docenten met weinig tijd om proefschriften te lezen. Niet alleen docenten Nederlands, maar ook zeker (of misschien wel: vooral) docenten van andere vakken.

Bij het vak Nederlands leren leerlingen een goede tekst te schrijven. Het gaat dan om *generiek* leren schrijven. Aspecten die docenten Nederlands belichten zijn bijvoorbeeld het bepalen van tekstdoelen, het aanbrengen van structuur in een tekst en het duidelijk formuleren van wat je bedoelt. Centraal staat *communiceren*: kun je een duidelijk verhaal overbrengen aan je lezer? Het is voor docenten Nederlands een kluit om ervoor te zorgen dat leerlingen de school *geletterd* verlaten: het niveau van leerlingen blijft dalen (OECD, 2018), terwijl de samenleving een steeds hoger taalniveau lijkt te vereisen.

Bij andere vakken dan Nederlands wordt ook veel geschreven. Docenten van andere vakken doen graag een beroep op de generieke schrijfvaardigheid van leerlingen en veronderstellen deze vaardigheden bij leerlingen veelal bekend. Er wordt dus bij andere vakken wel geschreven, maar niet zo zeer aan schrijfvaardigheid gewerkt. In feite is dat een gemiste kans. Bovendien lijken docenten niet erg tevreden over het niveau waarop leerlingen schrijven. Gezien de veelgehoorde klacht dat leerlingen "zo slecht formuleren", lijkt de noodzaak er dus wel te zijn om aandacht te besteden aan schrijven bij andere vakken.

Tot slot noemen we nog een belangrijke reden om er niet zomaar van uit te gaan dat er dus niets gedaan hoeft te worden aan de ontwikkeling van schrijfvaardigheid bij andere vakken. Naast generieke vaardigheden, die inderdaad vanuit het vak Nederlands meegenomen kunnen worden naar

andere vakken, spreekt de docent met een opdracht voor zijn vak ook nog een andere schrijfvaardigheid aan: de *vakspecifieke schrijfvaardigheid*.

Om te beginnen kent elk vak kent zijn eigen tekstgenres. Soms worden dezelfde termen genoemd bij verschillende vakken, maar dat wil niet zeggen dat de termen dan ook dezelfde invulling kennen. Neem bijvoorbeeld het "betooft". Een betooft bij Nederlands is een persuasieve tekst, die bedoeld is om een lezer te overtuigen. De tekst is dan ook heel publieksgericht. De schrijver gebruikt argumenten en retorische middelen die passen bij zijn publiek, om zijn betooft kracht bij te zetten. De schrijver maakt ook gebruik van informatie om zijn argumenten te ondersteunen, maar kan die, omwille van zijn retorische doel, selectief inzetten. Een betooft bij geschiedenis is niet zo zeer gericht op het publiek. Er wordt vaak wel een mening gegeven, maar deze mening is een zorgvuldig afgewogen oordeel op basis van gedegen onderzoek naar de historische context en perspectieven die er zijn omtrent het besproken onderwerp. Het oordeel is gebaseerd op historische bronnen, die zorgvuldig geëvalueerd worden op hun betrouwbaarheid.

Het is heel voorstelbaar dat een leerling die bij geschiedenis wordt gevraagd om een betooft te schrijven, zijn opvatting van een betooft als "persuasieve tekst" meeneemt. Zeker als de docent geschiedenis ook nog eens accentueert dat de kennis die vanuit Nederlands is geleerd, bij deze opdracht dient te worden toegepast. Er zijn natuurlijk ook zeker raakvlakken tussen de twee soorten betogen. Er is een grote overlap aan criteria, zoals een duidelijke basisstructuur en coherentie. Maar er zijn ook vakspecifieke criteria, zoals het beoordelen van bronnen op betrouwbaarheid en representativiteit, het wegen van argumentatie vanuit verschillende perspectieven, inhoudelijke juistheid en correct gebruik van begrippen. De vakspecifieke aspecten van het betooft zullen door de geschiedenisdocent wel eerst duidelijk gemaakt moeten worden, wil de leerlingen zich aangesproken voelen in zijn vakspecifieke schrijfvaardigheid.

Twee zaken zijn hierbij van belang. Ten eerste moeten een leerling weten wat er precies van hem verwacht wordt bij een bepaalde schrijftaak. Hoe ziet de tekst eruit, waar moet de tekst aan voldoen? Daarnaast is het schrijfproces van belang; wat moet een leerling doen om zo'n tekst goed te schrijven?

Wat we kunnen bereiken als de vakdocent in zijn lessen ingaat op bovenstaande twee aspecten, kunnen we laten zien aan de hand van een voorbeeldleerling die meedeed aan ons onderzoek bij het vak geschiedenis. De leerling kreeg voorafgaand aan het onderzoek een geschiedenisstaak voorgeschoteld. Deze taak bevatte de vraag: *In hoeverre vind jij dat westerse musea koloniaal*

erfgoed moeten teruggeven aan landen van herkomst? Er werden vier korte bronnen bij gegeven. Een eerste bron was een relaas van Multatuli, waarin hij aankaarte hoe er werd omgegaan met de bewoners van Indonesië ten tijde van de Nederlandse bezetting. De tweede bron was afkomstig uit een verslag van de Nederlandse kunstenaar Wijnand Otto Nieuwenkamp, die meeding op en militaire expeditie om kunstvoorwerpen "veilig te stellen". De derde en vierde bron waren actuele nieuwsberichten. Een ervan was een interview met een Balinese vorst, die zei kunststukken het liefst te willen gebruiken bij rituelen, in plaats van ze alleen maar tentoon te stellen in musea. De laatste bron was een weergave van een interview met de directeur van het Museum Nasional in Indonesië. Hij gaf aan dat het teruggeven van objecten ook kan leiden tot spanningen, want van wie is het object precies? Ook noemde hij niet alle erfgoed even waardevol. De gehele opdracht inclusief bronnen besloeg twee pagina's tekst.

Zonder verdere instructie schreef de leerling uit 5 vwo het volgende:

Ik denk dat westerse landen een deel van het koloniaal erfgoed zouden moeten teruggeven. Soms is het alleen moeilijk om erachter te komen waar iets oorspronkelijk vandaan komt en soms kan niet verzekerd worden dat het ook weer in een museum terecht komt, in dat geval zou het waarschijnlijk beter zijn als het in een westers museum blijft. Volgens mij is het vooral belangrijk dat het voor iedereen (in een museum bijvoorbeeld) goed zichtbaar is en, zo mogelijk, ook duidelijk gemaakt wordt waar het oorspronkelijk vandaan komt.

De tekst bevat een duidelijke mening van de leerling, maar die mening is onvoldoende onderbouwd. De bronnen lijken weinig invloed te hebben gehad op de mening van de leerling; ze worden zelfs niet eens genoemd. Het is alsof de leerling de bronnen niet of nauwelijks heeft gelezen. Het is dus geen zorgvuldig afgewogen oordeel op basis van historisch onderzoek en de tekst draagt daarom niet bij aan de gehele discussie over koloniaal erfgoed waarnaar de opdracht verwijst.

Toegegeven, de taak gaf ook geen richting in hoe de uiteindelijke tekst eruit zou moeten komen te zien. De enige aanwijzing die de leerling kreeg was om gebruik te maken van de gegeven bronnen. Er werden verder geen aanwijzingen gegeven over tekstlengte of elementen die in de tekst aanwezig moesten zijn. In feite zitten de vereiste elementen wel in de vraag: een "in hoeverre-vraag" vereist logischerwijs een genuanceerd antwoord, met argumenten. Dit is echter impliciet, terwijl leerlingen vanuit werkboeken en examen-vragen wel gewend zijn om meer gestuurd te worden richting het gewenste antwoord. Veel vragen in werkboeken worden uitgesplitst in stappen a-b-c,

of vragen naar argumentatie vanuit slechts één perspectief. Toch hebben we in deze studie gebruik gemaakt van open en overkoepelende vragen, juist omdat deze vragen de leerling stimuleren de kwestie *als geheel* te beschouwen.

Na afloop van een korte lessenreeks schreef dezelfde leerling een tekst naar aanleiding van de vraag: *In hoeverre is de term 'politioele actie' achteraf bezien te rechtvaardigen?* Ook hier waren weer vier korte bronnen gepresenteerd. De eerste bron was een fragment uit een radiotoespraak die Luitenant Gouverneur-Generaal Van Mook gaf, waarin hij de redenen voor de politioele acties uitlegt. In de tweede bron maakt de hoofdredacteur van Vrij Nederland zich in een opiniestuk uit 1947 boos over de politioele acties; hij noemt ze "oorlog". In een derde bron uit 1969 vertelt Joop Hueting, oud-soldaat in Nederlands-Indië tijdens de politioele acties, over het massa-geweld in Indonesië in een uitgeschreven tv-interview. In een laatste bron uit 2005 reflecteert een oud-soldaat terug op zijn tijd in Nederlands-Indië, in het kader van onderzoek door het Verzetsmuseum. Net als in de voormeting besloeg de gehele opdracht ook nu weer twee pagina's tekst. Na instructie over het schrijven van teksten bij geschiedenis schreef de leerling de volgende tekst:

Toen Soekarno in 1945 na de Japanse capitulatie de onafhankelijke Republiek Indonesië uitriep, probeerde Nederland de koloniale situatie te herstellen. Om dit te bereiken trad Nederland met veel geweld op in twee "politioele acties". Tegenwoordig is de vraag in hoeverre deze term te rechtvaardigen is.

In 1947 zei Luitenant Gouverneur-Generaal Van Mook dat er geen andere mogelijkheid was om rust en orde te verkrijgen en uiteindelijk de vrije Verenigde Staten van Indonesië op te richten. In een artikel van het dagblad De Nieuwe Courant staat ook dat de Nederlandse regering duidelijk wilde vaststellen dat deze maatregelen niet tegen het Indonesische volk maar de machthebbers en groepen die de werkelijke bevrijding van Nederlands-Indië in de weg staan.

Henk van Randwijk publiceerde in juli 1947 in Vrij Nederland een commentaar waarin hij stelde dat het geen politioele acties zijn, maar oorlog. Hij maakt in zijn commentaar ook vergelijkingen met het verkeerd begrepen nationaal belang van het Duitse volk in de tweede wereldoorlog.

In 1969 vertelde Joop Hueting, een oud-soldaat in Nederlands-Indië, in een tv-interview van oorlogsmisdaden die hij de inlichtingendienst heeft zien verrichten. Hij maakt duidelijk dat het geen incidentele gevallen waren, maar de normale gang van zaken.

De Nederlandse soldaat Jan Kuiper die zich meldde als oorlogsvrijwilliger voor de bevrijding van Indië vertelde in een gesprek met een documentalist van het Verzetsmuseum dat hij in loop der tijd langzaam twijfels over de militaire aanwezigheid van Nederland kreeg.

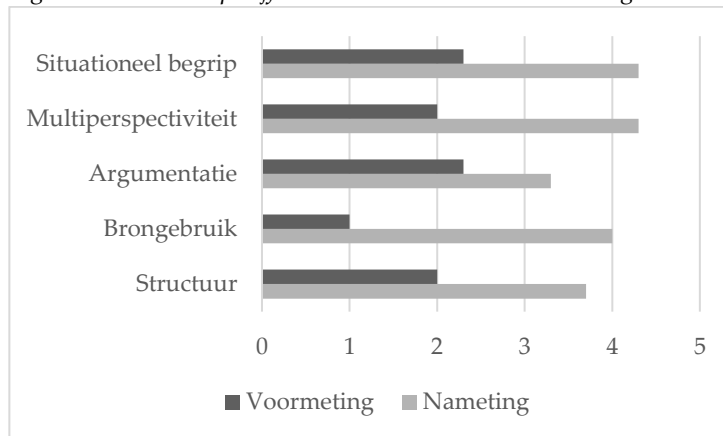
Toen leek het voor sommige mensen misschien rechtvaardig wat er in Nederlands-Indië gebeurde, maar ondertussen weten wij dat het niet om politioele acties, maar om oorlogsvoering ging, met het doel de onafhankelijkheid van Indonesië te stoppen.

De tekst van de leerling neemt duidelijk toe in kwaliteit. Scores van een onafhankelijke jury van geschiedenisdocenten bevestigden dit. Deze jury kende punten toe aan de hand van een schaal met vijf voorbeeldteksten, waarbij de

minst goede tekst een fictieve score had van 70 punten, de middelste tekst 100 punten, en de beste tekst een score had gekregen van 130 punten. De eerste tekst van de voorbeeldleerling kreeg een holistische score van 83 punten; de tweede tekst kreeg 116 punten. Waar de eerste tekst dus ondergemiddeld scoorde, scoorde de tweede tekst bovengemiddeld.

De jury gebruikte ook nog een rubric met een schaal van 1 (zwak) tot 5 (goed) met vijf aspecten die van belang zijn in historische teksten (ontleend aan De La Paz et al., 2017). Ook hier (Figuur P1) zien we een duidelijke stijging in scores:

Figuur P1. Scores op vijf criteria voor de voorbeeldleerling



Voor het vak filosofie kunnen we eveneens een voorbeeld geven. Naar aanleiding van de vraag *In hoeverre is de mens in staat tot zuiver altruïsme?*, met daarbij vier bronteksten, schreef een leerling uit 4 vwo zonder instructie de volgende tekst:

Zuiver altruïsme, het belangeloos helpen van anderen zonder iets terug te verwachten, bestaat dat eigenlijk wel? Volgens hoogleraar sociale filosofie Roos Vonk zit er achter een onbaatzuchtige daad vaak een egoïstische reden en zijn de twee concepten niet zo tegenovergesteld als vele denken. Vonk suggereert dat er allerlei situaties zijn die op het eerste oog alleen altruïstische motieven lijken te hebben, waarbij later blijkt dat er ook egoïstische redenen achter kunnen verschuilen. Denk bijvoorbeeld aan een situatie waarbij jij iets voor iemand doet en verwacht dat zij iets voor jou terug zullen doen.

Volgens evolutiebioloog Richard Dawkins kan het menselijke moraal worden verklaard door onze 'zelfzuchtige' genen. Door de evolutietheorie zijn mensen niet per se gericht op zichzelf, maar om onze genen in stand te houden. Wetenschapsjournalist Matt Ridley is het eens met Dawkins en noemt het voorbeeld van geld nalaten aan je kinderen. Volledig egoïstische mensen zouden dit niet doen maar Dawkins zijn theorie zou dit wel kunnen verklaren.

Moraalfilosoof Adam Smith en boeddhistische monnik Matthieu Ricard zeggen daarentegen dat er wel zuivere altruïstische beslissingen gemaakt kunnen worden door de mens. Smith gebruikt hierbij het voorbeeld van emoties die van persoon op ander persoon kunnen overslaan. Ricard zegt zelfs dat kinderen van nature al onbaatzuchtig geboren zijn.

Wat we in deze tekst voornamelijk zien, en wat veel andere leerlingen ook deden, is dat de bronnen worden samengevat in de tekst. Hoewel de leerling dat in feite goed doet, is het niet wat een docent filosofie vraagt van een leerling met de gegeven opdracht. De gestelde filosofische vraag verlangt nog een *afweging* van de gegeven perspectieven. Het *eigen* filosofisch denken van de leerling is niet zichtbaar in deze tekst.

Na de lessenreeks voor het schrijven van filosofische teksten, kreeg dezelfde leerling een nieuwe schrijftaak, met de stelling: *Docenten op Het Hoogt* zouden minder verplichtingen moeten opleggen en meer uit moeten gaan van eigen verantwoordelijkheid van leerlingen*, waarbij opnieuw vier bronnen werden gepresenteerd. Dezelfde leerling schreef de volgende tekst:

Minder verplichtingen voor kinderen?

Op de gemiddelde school, en ook Het Hoogt worden jongeren ontzettend veel verplichtingen opgelegd in de vorm van regels of normen. Leerlingen zijn het niet altijd eens met deze verplichtingen en weten soms niet eens of ze überhaupt nuttig zijn. Dit roept de volgende vraag op: Moeten docenten op Het Hoogt minder verplichtingen opleggen en meer uit moeten gaan van eigen verantwoordelijkheid?

Als we over vrijheid en verplichtingen praten moeten we eerst een onderscheid maken tussen twee belangrijke vormen van vrijheid: positieve vrijheid en negatieve vrijheid. Volgens Berlin is er sprake van positieve vrijheid wanneer iemand volledig meester over zichzelf is en zijn eigen keuzes welbewust maakt. Volgens hem kan paternalisme, ook uit goede bedoelingen, gevaarlijk zijn voor iemand en zou daarom beargumenteren dat minder verplichtingen op Het Hoogt een goed idee zou kunnen zijn.

Volgens sommige filosofen, zoals John Stuart Mill, is paternalisme iets dat zo min mogelijk zou moeten worden gebruikt. Desalniettemin, als we met behulp van maatregelen het schaden van mensen kunnen voorkomen mogen we dit doen. Op Het Hoogt kunnen in sommige gevallen verplichtingen en goede impact hebben; zoals het verplichten van brugklaskinderen om op het schoolplein te blijven. Hier wordt hun vrijheid afgenomen om te zorgen dat ze niet verongelukken tijdens schooltijd. Bovendien kunnen er misschien leerlingen bij zitten die niet in staat om autonome beslissingen te maken omdat ze nog te jong zijn.

Wetenschappelijk gezien is de prefrontale cortex, het deel in je brein dat bepalend is voor het maken van verstandige keuzes nog lang niet volgroeid voor kinderen op een middelbare school. Volgens Eveline Crone is die pas klaar met ontwikkelen wanneer iemand 24 is. Je zou hiermee kunnen beargumenteren dat leerlingen niet zelfstandig genoeg zijn om sommige verplichtingen te verwijderen, maar volgens de wet is een persoon vanaf 18 jaar wel verstandig en zelfstandig genoeg om (bijna) alles zelf te kunnen doen dus dit hoeft niet per se te kloppen.

* "Het Hoogt" is een fictieve school. Oorspronkelijk stond hier de echte naam van de school van de leerlingen die de schrijftaak maakten. Leerlingen pasten de vraag dus toe op hun eigen school-situatie.

Een argument voor deze vraag zou kunnen worden gegeven door Michel Foucault die zegt dat onze vrijheid wordt beperkt door instituties zoals scholen. Volgens Foucault zouden verplichtingen niet eens per se nodig zijn omdat de leerlingen zelf elkaar in de gaten houden.

Concluderend, ik vind persoonlijk dat minder verplichtingen op Het Hoogt niet zozeer nodig zijn omdat ik het eens ben met de woorden van Eveline Crone, het brein van een puber is nog lang niet volledig volgroeid en daarom zou te veel vrijheid voor onwijze keuzes kunnen zorgen.

Terwijl deze opdracht en bronnen als complexer werden bestempeld door een onafhankelijke jury, schreef de leerling een betere tekst dan bij de eerste taak. In deze tekst laat de leerling zien dat hij zelf kan nadenken over de informatie uit de bronteksten. De verschillende perspectieven uit de bronnen worden niet alleen op een rij gezet, maar direct gekoppeld aan de gegeven casus van de eigen school.

SCHRIJFLESSEN

Hoe bereiken we dit? Om te beginnen ontwierpen de vakdocenten die meededen aan ons onderzoek zelf een aantal nieuwe schrijftaken, op basis van ontwerpprincipes voor effectieve leer-schrijftaken, afgeleid uit wetenschappelijk onderzoek:

Ontwerpprincipes voor schrijf-leertaken

1. Ontwerp schrijftaken die evaluatief van aard zijn. Stel dus een vraag die een afweging oproept, die de leerling baseert op bronnenonderzoek. Dit helpt leerlingen om een kwestie als geheel te bekijken en beoordelen.
 2. Voorzie de taak van bronnen (tekstueel of visueel, of beide), waarin verschillende perspectieven worden gerepresenteerd. Bronnen spreken elkaar dus tegen, of ze vullen elkaar aan. Wanneer er twee vaardigheden, lezen én schrijven, worden aangesproken, stimuleert dat het leerproces. Primaire bronnen hebben de voorkeur; leerlingen hebben bij primaire bronnen meer aandacht voor bronkenmerken, zoals: wie is de auteur precies? Vooral voor het vak geschiedenis is dat van belang.
 3. Zorg ervoor dat de taak voor leerlingen binnen één lesuur te maken is. Op deze manier is het schrijfproces voor jou als docent zichtbaar en dat geeft je de mogelijkheid om het proces van leerlingen te begeleiden. Bak de kwestie dus goed af in de vraagstelling en presenteer maximaal vier bronnen. Dit resulteert in teksten van 200-300 woorden, die ook nog behapbaar zijn als het gaat om lees-/nakijkwerk.
-

Verder ontwierpen we op basis van wat we weten uit eerder schrijfonderzoek een aantal lessen die inzetten op de twee eerder genoemde aspecten: een beeld vormen van de taakeisen en instructie over het lees-schrijfproces. Vakdocenten voerden de lessenreeks uit in 4, 5 en 6 vwo op verschillende scholen, bij

zowel bij geschiedenis als bij filosofie, waarbij de lessenreeks steeds een vakspecifieke invulling kreeg.

Overzicht van de lessenreeks

Les 1: alle leerlingen maken een schrijftaak over het vak. Het "ervaren" van de taak is nuttig om te kunnen reflecteren. Wat is moeilijk aan deze taak? Zo kunnen leerlingen hun recente ervaring gebruiken bij het bespreken wat er moeilijk is aan het schrijven van teksten voor dat vak en het schrijven van teksten in het algemeen. Schrijven op de computer verdient de voorkeur boven werken met pen en papier bij het schrijven van langere teksten; op de computer gaat schrijven sneller en je kunt makkelijker reviseren. Doordat er minder aandacht uitgaat naar het motorische, kan er meer aandacht uitgaan naar inhoud en structuur.

Les 2: leerlingen reflecteren op hun eigen proces en krijgen vervolgens een nieuwe vakspecifieke strategie aangeboden. Deze bestaat voor beide vakken uit zeven stappen, verspreid over drie fasen: *lezen*, *denken* en *schrijven*. Daarnaast staat *monitoren* centraal. De docent benadrukt vakspecifieke elementen in de instructie. Dat kan lastig zijn, omdat je je als vakdocent misschien niet altijd bewust bent van vakspecificiteit. In gesprek gaan met collega-docenten van andere vakken, of juist van je eigen vak, over wat schrijven bij jouw vak inhoudt, kan daarbij helpen. Vervolgens is het van belang de nieuwe strategie voor te doen, ofwel te "modelen". In ons onderzoek maakten we gebruik van een alternatief: we maakten een video waarin leerlingen te zien zijn ("modeling peers") die de strategieën toepassen. De video nu en dan pauzeren is aan te bevelen om te kunnen stilstaan bij wat te zien is. Wat doet de voorbeeldleerling? Is dat handig? Doe jij dat ook zo? De QR-codes verwijzen je naar de video's die we maakten voor elk vak.

Geschiedenis



Filosofie



Les 3: In een volgende les krijgen de leerlingen een aantal voorbeeldteksten te lezen van klasgenoten. Deze teksten geven aanleiding om te discussiëren over de goede en zwakke punten van elke tekst, om vervolgens gezamenlijk te komen tot een lijstje van circa vijf criteria voor goede *historische* dan wel *filosofische* teksten. Ook hier blijf je uitgaan van jouw vak. Zo belicht je aspecten die vakspecifiek zijn.

Les 4: Daarna gaan de leerlingen aan de slag met een nieuwe schrijfofdracht. Deze is bij voorkeur in één lesuur te maken, zodat de docent de leerlingen kan ondersteunen bij het schrijfproces. Het is in ieder geval van belang dat er geschreven wordt in de les zelf, zodat het proces van de leerlingen voor jou als docent ook zichtbaar is. De geleerde strategie is het uitgangspunt bij het geven van feedback op het lees-schrijfproces. Als een leerling vastloopt, kun je bijvoorbeeld vragen bij welke stap de leerling vastliep, en waarom.

Les 5: De docent bespreekt in een vervolgles een of meer teksten van leerlingen in de les. Daarvoor kan gebruik gemaakt worden van een van de feedbackopties (zie kader "Hoe bespreek je een schrijftaak?"). Steeds wordt er teruggesproken naar de strategie. Wat moet er gebeuren in het proces, om te komen tot een beter proces? Welke stap uit de strategie verdient aandacht? Deze les draait om efficiënte feedback; effectief voor alle leerlingen, maar niet te arbeidsintensief voor de docent. Nakijken van dertig schrijfofdrachten is immers erg veel werk.

Les 6/7: Herhaling van les 4 en 5, met een nieuwe schrijfofdracht. Waarschijnlijk kan de leerling met minder hulp af en wordt de feedback steeds gericht.

Hoe bespreek je een schrijftaak?

De leerlingen hebben teksten geschreven en die bij jou ingeleverd. En nu? Je kunt de schrijftaken in een vervolgles bespreken, bijvoorbeeld op één van onderstaande manieren. Je maakt bij deze besprekingen gebruik van leerlingteksten die je anonimiseert. Een alternatief is om geanonimiseerde teksten uit een andere klas of ander leerjaar te gebruiken. **Richt je bij de bespreking (naast op inhoud en structuur) steeds óók op het proces: hoe kan een leerling tot een beter resultaat komen?**

Het maken van een tekstschaal

Kies drie teksten van verschillend niveau. Je kunt dat doen door de teksten globaal te bekijken en drie stapeltjes te maken: zwakke teksten, gemiddelde teksten en sterke teksten. Kies er vervolgens één uit elke stapel.

In de les laat je leerlingen (in duo's of groepjes) de teksten op volgorde leggen van niveau. Ze gaan de drie of vijf teksten dus rangordenen. Vervolgens bespreek je de rangordes klassikaal: waarom is de beste tekst dan het best? En wat ontbreekt er aan de zwakke tekst? *Welke stap in het proces zou beter uitgevoerd moeten worden voor een beter resultaat?*

Eigen teksten vergelijken met een tekstschaal

Kies drie teksten van verschillend niveau (zwak-gemiddeld-sterk). Presenteer ze op het bord. Geef aan wat je zwak-redelijk-sterk vindt aan de teksten. Laat leerlingen in duo's hun teksten uitwisselen en in jouw schaal plaatsen. Waar zou hun tekst in de schaal passen? Waar ligt dat aan? *Welke stap uit het stappenplan verdient blijkbaar meer aandacht?*

Een tekst van een willekeurige leerling bespreken

Vraag welke leerling zijn tekst op het bord besproken zou willen hebben. Laat de tekst van deze leerling op het bord zien en bespreek deze klassikaal. Wat is er goed aan, wat is er minder goed aan? *Wat kan de leerling doen om zijn tekst te verbeteren?*

Een goed voorbeeld presenteren

Kies een heel goede tekst uit en presenteer deze op het bord. Een goede voorbeeldtekst kan zwakkere leerlingen helpen om te zien hoe een goed eindresultaat eruit zou kunnen zien. Bespreek waarom je deze tekst zo goed vindt. Welke elementen maken de tekst zo sterk? *Hoe kom je tot die elementen, welke stappen moet je daarvoor zetten?*

Op criteria beoordelen

Formuleer eerst criteria voor goede vakteksten. Presenteer vervolgens een of meer teksten in de klas en leg ze naast de criteria. Aan welke criteria voldoet de tekst wel of niet?

Dit kan ook individueel gebeuren: laat leerlingen hun eigen tekst naast de criteria leggen. Of laat leerlingen de tekst van een medeleerling scoren op alle criteria. *Aan welke stap uit het stappenplan moet je nog aandacht besteden?*

DE RATIONALE ACHTER DE SCHRIJFLESSEN

Waarom zouden we leerlingen überhaupt schrijftaken geven in vaklessen? Vanzelfsprekend kunnen leerlingen hiermee laten zien wat ze in huis hebben aan vakspecifieke kennis en vaardigheden. Zo weten we waar ze staan in hun kennis- en vaardighedenontwikkeling. Daarnaast leren leerlingen door middel van schrijftaken communiceren. In het algemeen, maar leerlingen leren ook wat de verschillende vakken verlangen in teksten voor die vakken. Hoe historici schrijven, leer je bij geschiedenis. En hoe filosofen redeneren op papier, leer je bij filosofie.

Een andere belangrijke reden om schrijftaken in je vak te gebruiken is dat je door te schrijven je gedachten kunt ontwikkelen. Je kunt schrijvend leren: schrijven bevordert kennis of begrip van vakinhoud (Graham et al., 2020). Ook bevordert schrijven lezen (Graham & Hebert, 2010). Door te schrijven over wat je leest, word je gedwongen om hierover na te denken. Tot slot weten we dat schrijven en lezen geïntegreerd sterker werken dan apart (Graham et al., 2018).

Schrijven leidt niet altijd tot leren. Soms is een genre niet zo geschikt voor 'diep leren': maak je aantekeningen bijvoorbeeld, dan maak je niet zo snel nieuwe connecties. Bij een samenvatting of een verslag geldt hetzelfde. Over het algemeen kunnen we zeggen dat er beter begrip ontstaat van de leerstof met analysetaken, zoals essayschrijven.

Een andere situatie waarin schrijvend leren niet opgaat, is als tekstkwaliteit en leren elkaar tegenwerken. Het maken van een schrijfplan of outline kan bijvoorbeeld leiden tot een heel goed gestructureerde tekst, maar tot minder ontdekkingen, of zoals Baaijen en Galbraith (2018) het noemen: discovery. Aan de andere kant kan die discovery leiden tot een heel ongeorganiseerde tekst, omdat de leerling moeite krijgt om zijn gedachten op een rijtje te krijgen. Dan zal een leerling meerdere versies moeten schrijven van zijn tekst. Al met al zal er in schrijfopdrachten een balans moeten zijn tussen het ruimte geven aan discovery aan de ene kant, en het beteugelen van discovery aan de andere kant, omwille van communicatieve doelen. Een tekst moet uiteindelijk immers wel leesbaar zijn. Het is voor vakdocenten met name van belang dat zij zich bewust zijn van het doel van hun schrijfopdrachten. Waarom laat je leerlingen schrijven, met welk leerdoel?

Vakspecifiek schrijven

Schrijven bij geschiedenis is iets anders dan schrijven bij Nederlands of biologie, of welk vak dan ook. Natuurlijk zijn er kenmerken van goede teksten te

noemen die generiek zijn. We zien allemaal graag een goede structuur, alinea-indeling, kop-romp-staart. En ook zien we het liefst teksten waarin aandacht besteed is aan het taalgebruik, de spelling en de formulering. Toch kent elk vakgebied eigen vakspecifieke elementen in teksten, waar je niet aan voorbij kunt gaan (Carter, 2007). Bij geschiedenis wordt geschreven op basis van historische bronnen. De nadruk ligt op het kritisch bekijken van die bronnen; zijn ze betrouwbaar, representatief? Bronmateriaal kan op deze manier ingezet worden als ondersteuning bij argumentatie.

Bij filosofie gaat het schrijven ook nadrukkelijk om het proces. Het meest bekende genre in dit vak is het *filosofisch essay*: het schrijven zelf is een zoektocht naar het antwoord op een filosofische vraag.

Elk vak kent vakspecifieke invullingen van genres. Onderzoek van Coffin (2006) naar tekstfuncties bij geschiedenis resulteerde in de definitie van drie verschillende genrefamilies van geschiedenisteksten: "vastleggen", het optekenen van gebeurtenissen op een zo feitelijk mogelijke wijze; "uitleggen", waarbij oorzaken en gevolgen bijeengebracht worden, en "argumenteren", waarbij een leerling een kwestie van verschillende kanten bekijkt en zichzelf verhoudt tot een kwestie. Bij elk genre hebben we andere verwachtingen over hoe de tekst eruit moet zien. Bovendien heeft elk genre weer andere leereffecten. Hoewel dit onderzoek specifiek over het vak geschiedenis ging, veronderstellen we dat dit ook geldt voor andere vakken.

Taakeffectstudies bij geschiedenis (Voss & Wiley, 1997; Van Drie et al., 2006; Newell & Winograd, 1995) wijzen uit dat de leereffecten het grootst zijn bij een argumentatieve taak. In een studie van Van Drie (et al., 2006) werden twee soorten opdrachten vergeleken. In de ene onderzoeksconditie beantwoordden leerlingen de vraag: *Hoe kun je de veranderingen in het gedrag van jongeren in de jaren zestig verklaren?* Een verklarende vraag. In de andere onderzoeksconditie gingen leerlingen aan de slag met de vraag: *Waren de veranderingen in het gedrag van jongeren in de jaren zestig revolutionair?* Een evaluatieve vraag, waarbij wordt gevraagd naar een eindoordeel op basis van historische bronnen. Deze laatste vraag had een groter leereffect: de historische redeneringen in de leerlingteksten waren beter.

Argumenteren is een term die soms verwarring oproept. Het onderzoek van Coffin onderscheidt onder "argumenteren" weer drie verschillende termen: *een standpunt verdedigen*, *een kwestie beschouwen* en *een standpunt betwisten*. Elk van deze genres vereist een eigen structuur. Een ander onderscheid is het enkel- versus meervoudig perspectief (Monte-Sano & Allen, 2018): is de argumenterende tekst er een waarin één kant wordt beargumenteerd, of wordt een

kwestie van verschillende kanten bekeken? In het onderwijs worden deze termen nog wel eens door elkaar gebruikt, of de overkoepelende term argumenteren wordt gebruikt, waarbij de precieze verwachtingen niet expliciet worden gemaakt, terwijl dat wel belangrijk is. Is de instructie vaag, dan neigen leerlingen naar het schrijven van een eenzijdige tekst (Monte-Sano & Allen, 2018). Het is dus van belang om in een schrijftaak of in je les te expliciteren wat je precies voor tekst verwacht: je werkt daarmee aan een goede beeldvorming van de taakvereisten.

Schrijven in de vakken is dus erg nuttig: het kan het leren bevorderen. Gebruik je een schrijftaak in de les, dan is het noodzakelijk om na te denken over welk tekstgenre je verwacht en waarom. Tot slot kent elk vak zijn eigen vakspecifieke tekstgenres en -elementen, die aandacht verdienen in de vakles. Onze ontwerpprincipes voor schrijftaken helpen je op weg om nuttige leer-schrijftaken voor vaklessen te ontwikkelen. Je werkt met zulke taken aan taalvaardigheid en vakvaardigheid tegelijk.

Schrijfprocessen begeleiden

De schrijflessen uit onze onderzoeken zijn gebaseerd op onderzoek naar het ontwikkelen van algemene schrijfvaardigheid. Hiernaar is namelijk veel onderzoek gedaan. We weten uit eerder onderzoek dat het zinnig is om aandacht te besteden aan het schrijfproces van leerlingen, en om daarbij rekening te houden met hun schrijfvoorkeuren.

Een schrijfstrategie kun je goed onderwijzen met behulp van strategie-instructie aan de hand van het model van Self-Regulated Strategy Development (SRSD) (Harris & Graham, 2009): grof gezegd wordt een nieuwe strategie volgens dit model eerst uitgelegd, vervolgens voorgedaan, en tot slot inge oefend.

Het ontwikkelen van een strategie die geschikt is voor de taak en het vak vereist bewustzijn. Om te komen tot een strategie die je kunt uitleggen aan leerlingen, moet je je eerst bewust zijn van het proces dat je eigenlijk van leerlingen verwacht. In ons onderzoek zijn de twee strategieën voortgekomen uit onderzoek naar generieke lees-schrijfstrategieën, gecombineerd met vakspecifieke lees- en schrijfstrategieën. Dit resulteerde per vak in één strategie, die toegespitst was op het schrijfgenre. Om bij elke taak dus schrijfprocessen te kunnen begeleiden, is het noodzakelijk om na te denken over de vraag: welk proces verwacht ik van leerlingen bij deze taak? Het kan ook juist taakontwikkeling stimuleren, door deze vraag éérs te stellen. Welke processen wil ik leerlingen leren, en welke taak stimuleert precies *deze* processen?

CONCLUSIE

Al met al beveel ik met deze proloog aan vakdocenten aan om aandacht te besteden aan het ontwerpen van schrijftaken passend bij de vakleerdoelen en het geven van vakspecifieke schrijfprocesinstructie. Ik hoop dat de hier beschreven lessen hierbij een inspiratie zijn voor vakdocenten om schrijfinstructie in hun les te integreren.

CHAPTER 1

GENERAL INTRODUCTION

Nowadays, life is full of texts. Phones and computers confront us with a variety of text types, forcing us continuously to be readers. Over the past decades, increased literacy standards have heightened the need for advanced literacy skills. Our current society requires readers who can select, organize, and critically assess this constant stream of information. The ability to be an analytical and critical reader is quite challenging to obtain. Reading texts on screens has become prevalent, instead of reading paper newspapers or books. Screen texts elicit a different kind of reading process: screens encourage us to read fast and superficial. Digital reading thus challenges the ability to read deep.

The societal importance of writing skill development is also prominent. A high writing proficiency is a prerequisite for independent participation in social life. Furthermore, writing skills are important for students' cognitive development (Graham et al., 2020), and writing as well as writing education can also enhance students' reading skills (Graham & Hebert, 2011).

In school settings, students are repeatedly asked to read texts, mostly with the goal of knowledge construction. To process new knowledge, reading activities are habitually accompanied by writing. Note-taking, answering study questions, and writing reports or essays, are examples of such writing activities. In education, students are thus required to both read and write on an advanced level. Especially in upper secondary grades, where the texts to read or write are long, and contain complex, abstract content.

However, writing activities across school subjects are rarely intended to be opportunities to learn-to-write; most frequently, the aim of writing in content classes, is to assess students' discipline specific reasoning or to process content explained by the teacher. Writing is thus rarely utilized as a learning activity in the sense of knowledge development, although the act of writing is well suited to do so (Graham et al., 2020). Possibly, learning content and learning to write could go hand in hand. However, this potential has not fully been utilised.

Meanwhile, Dutch students' literacy proficiency seems to be declining, according to reports of PISA (OECD, 2018): an international comparative study that tests the skills and knowledge of 15-year-olds from about 80 different countries. This trend does not only apply to the Dutch context: international reports have also revealed a worrying decrease of students' ability to read and write critically (Goldman, 2012). With the current research project, we tried to tackle this literacy decrease.

First steps have been taken by educational researchers (e.g., Lee & Spratley, 2010; Moje, 2008; Shanahan & Shanahan, 2008), who have called for the development of students' literacy skills to be seen as a responsibility for all teachers, and not of the language teacher alone. The Dutch policy makers (Ministry of Education, Culture, and Science, 2009) have adopted this view. A similar development has occurred in the United States, with the adoption of "Standards for Literacy in History/Social Studies, Science, & Technical Subjects" in the Common Core State Standards (National Governors Association & Council of Chief State School Officers, 2010). With these standards, literacy is acknowledged to differ from subject to subject, and learning to communicate in a subject should therefore be part of the subject objectives (Shanahan & Shanahan, 2017). After all, language is inseparable from knowledge acquisition and knowledge exchange, and should therefore concern all teachers.

Literacy being a shared responsibility does not mean that all teachers are language teachers. Instead, content teachers might take responsibility for the aspects of literacy that are part of their discipline, and thus focus on disciplinary literacy. With this view prevailing among content teachers, literacy development will no longer be considered to be a burden, or even an exasperation; it will be seen as an essential part of a subject, to learn how to communicate knowledge and reasoning in that subject.

Nonetheless, the outdated view that literacy is a matter for the language teacher, is still prevailing in educational practice. Although many teachers state that they support shared responsibility, they do not act on their beliefs (Graham et al., 2014). With this thesis, we aimed to align teachers' practice with their (unobtrusive) beliefs about literacy, and government principles.

1. MAIN CONCEPTS

1.1 Disciplinary Literacy: Reading and Writing

For a long time, reading and writing have been considered generalist notions: teachers often have the idea that from the moment students master the basics

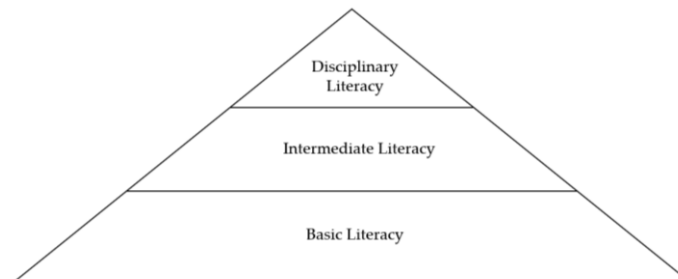
of reading and writing, they can successfully read and write anything. Indeed, the role of basic skills cannot be excluded: reading and writing highly depend on these generic skills. However, we can distinguish between multiple levels of literacy, disciplinary literacy being the most advanced form of literacy (Figure 1.1; Shanahan & Shanahan, 2008). Although Shanahan and Shanahan concentrated this model on reading, it could also be applied to writing. According to the model, basic literacy concerns highly generalizable skills needed for all or most reading and writing tasks. For example, basic decoding, handwriting, and spelling skills, understanding of conventions in texts, word comprehension, and basic fluency routines. The basic literacy level is to be achieved before secondary grades.

Support for students to develop literacy skills is drastically reduced when students reach proficiency in the basic level. However, it is a misconception that students need no support to master the more advanced literacy levels (intermediate and disciplinary literacy). Since these literacy levels are assumed to be reached at secondary school, secondary school teachers are likely to provide support and instruction, to reach the advanced levels.

Every discipline has its nuances, and with regard to text genres, these are the use of concepts, and heuristics. Goldman and colleagues (2016) clustered the differences which can be appointed between disciplines into five categories: (1) epistemology; (2) inquiry/reasoning strategies; (3) underlying concepts; (4) text genres; and (5) discourse/language character. Thus, addressing subject specificity in each of these five domains is an important task for the content teacher.

Within the Dutch education system, the context of our research, teaching writing skills is usually the responsibility of the language department, and not part of other subjects. Even for subjects with high demands in literacy, such as history and philosophy, learning to communicate disciplinary reasoning is no vital part of the curriculum. This might be due to the examination program: content knowledge is mostly examined by means of questions requiring short answers, and not by, for example, document-based essay questions. These would imply more writing, which again would reinforce the need for advanced writing skills. Apparently, short answers are not regarded to be "texts", which could be considered a problem.

Figure 1.1 *The Increasing Specialization of Literacy Development* (Shanahan & Shanahan, 2008, p. 44)



As a result of examination through question answering, focusing on content, subject objectives do not contain explicit literacy goals; these generally remain implicit. For example, the final attainment levels for philosophy do include argumentative skills, which potentially address students' speaking, reading, and writing skills. As a consequence, in most subjects, reading-writing tasks are mainly means for the teachers to assess students' thought processes. Consequently, teachers spend little time guiding students in their process, and hardly provide instruction on reading and writing (Gillespie et al., 2014; Motart et al., 2009). Overall, content teachers do not seem to add much to learning-to-write.

Nevertheless, using writing tasks as a learning activity can add value to school subjects. Writing itself can contribute to the learning process (Graham et al., 2020), as the writing process can allow thoughts to develop. After all, being forced to pause to formulate thoughts, can in turn lead to new thoughts (Klein, 1999). This idea is known as writing-to-learn.

A combination of reading and writing, source-based writing, could be even more stimulating (Graham, 2020). A meta-analysis of Graham and Hebert (2011) has shown that writing and writing instruction might enhance reading performance. Furthermore, a subsequent meta-analysis (Graham et al., 2018) showed that reading instruction improved students' writing as well. Reading and writing are thus intricately connected. Moreover, since writing about content has been shown to reliably enhance content learning (Graham et al., 2020), the combination of reading, writing, and learning seems judicious.

However, writing task design is a deliberate process, since writing does not necessarily imply learning; there are conditions attached. For instance, the text genre affects the learning effect (Applebee, 1984). Argumentative writing, for example, tends to create a deeper thought process and better understand-

ding of learning than summarizing, narrative writing or explaining (Wiley & Voss, 1999). In addition, it is important to consider subject-specific elements: characteristics of a good text differ from subject to subject and are embedded in what is common in those subjects (Bazerman, 1981). Overall, it can be quite a challenge for teachers to integrate effective subject-specific writing tasks into their subject teaching, and teaching aids are hardly available.

1.2 Writing Instruction and Support

When a writing task is assigned, research has shown it is effective to support students' writing process. How can a student learn to write a good text? What cognitive processes should be activated? From research conducted in the field of writing and writing instruction, we can derive recommendations for how to do this effectively (Graham & Harris, 2018). For example, by explicit writing strategy instruction or supporting students' writing process. This domain of research, however, is generally focused on learning-to-write, and not on content classrooms. Although writing instruction has been validated as effective for content classrooms as well (i.e., writing quality and knowledge gain) (De La Paz, 2005; Foxworth & Mason, 2017), subject teachers do not always have the knowledge to provide writing instruction or support, like language teachers have (Gillis, 2014; Moje, 2008).

In practice, when content teachers support students' writing, they tend to focus mostly on the product (Alkema, 2022; De Oliveira, 2011; Holdinga, 2013). Students often perform writing assignments out of sight of the teacher, for instance as homework, and students regularly only receive feedback on text quality after their texts are fully completed. This thus results in teachers providing feedback on the eventual texts students wrote, but not on how they came to this result.

To develop such process instruction and support, we will need understanding of cognitive processes needed to perform the particular source-based writing task. Several theoretical research has been conducted on cognitive processes of general reading and writing (Flower & Hayes, 1981), and the relationship between reading, writing, and learning (Galbraith, 1992, 2009; Graham, 2020; Hebert et al., 2013). However, disciplinary aspects of these processes have been less extensively researched. Studies researching disciplinary literacy are often based on comparisons of processes from experts with those of novices (e.g., Shanahan et al., 2011; Wineburg, 1991), or focus on functional linguistic analyses of the characteristics of disciplinary text genres (e.g., Coffin, 2006; Monte-Sano, 2010).

Up until now, the extent to which characteristics of expert reading and writing apply to subject-specific reading and writing of students in upper secondary grades, has been less widely researched. With our research, we will elaborate on studies on disciplinary writing at upper secondary level focusing on a single discipline (e.g., Van Driel et al., 2022b). In this thesis, we explore students' cognitive processes, in two different school subjects, by using a disciplinary lens. Furthermore, we transfer theoretical understanding of general reading and writing instruction into the practice of content classrooms, and thereby address issues concerning implementation of writing instruction, considering the conundrums mentioned earlier.

1.3 Innovating Practice

When innovating teachers' practice, an important influential factor is practicality. Writing instruction must be easily applicable to teachers who are not accustomed to providing such instruction because it is not part of the subject culture (O'Brien et al., 1995). Practicality should meet three criteria, all related to the ecology of the classroom and teachers' goal systems (Westbroek et al., 2020). A first criterion is instrumentality: procedures should be available, that show how the innovation is to be implemented. A second criterion is congruence: the innovation should be sufficiently congruent with regular practices and important goals to which the teacher is attached. And a third criterion is low cost: the estimated benefits of implementation should outweigh the effort required to implement the innovation. Keeping track of these practicality criteria is highly recommended for instructional design.

In our studies, congruence in particular turned out to be an important criterion for content teachers in their attitude towards writing instruction. As literacy development is not a priority on content teachers' agendas, it is even more important that we attempt to align literacy goals with learning goals of the subjects. Therefore, we consider flexibility of innovative designs to adapt to content goals and school contexts (curriculum, topics) as particularly relevant. We employed this criterion by involving teachers into the design process as much as possible.

Furthermore, innovation of teaching practices involves building teachers' knowledge and challenging their beliefs (Desimone, 2009). Simply designing new materials for teachers to implement, thus probably will not suffice to establish durable change; teachers' beliefs highly guide their teaching approaches. Therefore, in our studies, we increasingly aimed to challenge teachers' beliefs, to eventually initiate a lasting development of their writing

practice. In an initial implementation study we used a teaching manual only, in a subsequent study we added a professional development session explaining design principles to better prepare teachers for implementation; and in the final study we added several other guiding activities focused on reflection.

2. RESEARCH DESIGN AND SCOPE

The goal of this research was to design, implement, and test lesson materials for content teachers, in order to facilitate them in the development and support of secondary school students' disciplinary writing. Our main research question was:

Which instructional approach is suited to develop students' disciplinary literacy effectively and efficiently in upper secondary history and philosophy education?

We focused on two subjects: history and philosophy. Both of these subjects demand an elevated level of literacy: in upper secondary, pre-university levels, teachers expect students to read complex texts, to assess them critically, and to reflect on the content of source material in a well-written, coherent text. Such tasks transcend basic and intermediate literacy, and address disciplinary aspects.

In the field of history education, quite some research has been conducted to enhance students' writing. For example, De La Paz and colleagues (2010, 2017) and Van Drie and colleagues (2006, 2015) have explored how students' writing in history might be enhanced by interventions. We used these studies as a knowledge base for the design of our new lesson materials.

For philosophy, educational design studies have been scarce, and they rarely focus on writing. This was our rationale to start with the historical context; we built on previous intervention studies from history to eventually be able to transfer to another knowledge domain: the context of philosophy.

Regarding methodology, the overall research project can be characterized as design research. According to McKenney and Reeves (2019), educational design research can be typified by "its commitment to developing theoretical insights and practical solutions simultaneously, in real-world contexts, together with stakeholders" (p. 6). This methodology suited our goal of improving educational practice, in tandem with gaining theoretical knowledge of disciplinary reading-writing processes perfectly.

A typical design research, is an iterative and highly flexible organization of three core processes: (1) analysis and exploration, (2) design and construction, and (3) evaluation and reflection. In the first stage we explored the context of our studies: the educational context in two subjects, history and philosophy. We conducted background studies from two perspectives: teachers and students. Based on these background studies, we designed two instructional units (one for each subject) in the second phase. In the evaluation and reflection phase, we tested and evaluated our instructional units.

3. STRUCTURE OF THE THESIS

This thesis is a collection of four studies, on which we have reported in four chapters. These four chapters can also be read separately, since we have aimed to report the studies in four journal articles for publication. This created a certain amount of overlap between chapters.

Chapter 2 reports on a background study on students' perspectives on writing, while focusing on students' cognitive processes during source-based writing. In this study, 11th grade students performed writing tasks for history and philosophy while thinking aloud. The study provided the basis for designing the disciplinary strategy instruction, since it revealed the disciplinary aspects of the writing process. This study confirmed that the two subjects required different learning materials. We decided to continue with a design study for history first, and to make a transfer to philosophy at a later stage.

Chapter 3 presents a design study, which resulted in an instructional unit for historical writing, applicable to different topics. We started the research with a prestudy exploring the context from teachers' perspective. We conducted explorative interviews with history and philosophy teachers, to obtain a clear image of teachers' current knowledge of writing processes and instruction, and their habits around providing writing tasks and teaching writing. In real time, this prestudy preceded the think-aloud study presented in Chapter 2. However, we decided not to report on the prestudy in a separate chapter, but to integrate it in Chapter 3 (history teachers' perspective) and Chapter 5 (philosophy teachers' perspective) instead, for the benefit of coherence.

The prestudy, plus the think-aloud study from Chapter 2, as well as a literature search, provided the foundations for the instructional unit presented in Chapter 3. Overall, it resulted in two design principles, which formed the foundations for an instructional unit. After a trial study, in which the

instructional design was implemented and evaluated by two history teachers, recommendations for redesign were formulated.

In Chapter 4 we report on a larger-scale quasi-experimental study ($N = 268$) set-up in the history classroom, testing an optimized version of the instructional unit as presented in Chapter 3. The research design contained three experimental conditions, to assess the effectiveness of (a) implementation of writing tasks, developed by participating teachers, and (b) implementation of discipline-specific strategy instruction. In the first experimental condition, teachers developed writing tasks, which they used in their own history lessons. In the second experimental condition, participating teachers developed writing tasks tailored to their own year plans and they implemented additional writing strategy instruction. A third condition was a non-writing control condition.

In Chapter 5, we made the above-mentioned transfer to philosophy. The intervention materials we had designed for history, were adjusted to the discipline of philosophy, and subsequently implemented by three philosophy teachers. Again, in this study, teachers developed their own writing tasks, based on design principles. Additionally, in this study guiding activities were implemented, prompting teachers to reflect on their writing practice. We used qualitative measures to research teachers' interaction with the designed learning materials, and to explore if the intervention was effective for students' philosophical writing. Methods to measure effectiveness were not only the assessment of students' texts by independent jury teams, but also the teachers' reflections on their students' results.

In the general discussion, we discuss new insights concerning the main concepts presented in this introduction: disciplinary literacy, supporting writing, and innovation of teachers' practice. Furthermore, methodological issues, and directions for future research are discussed. We end with a series of recommendations for content teachers. As far as the subjects of philosophy and history are concerned, we will offer our readers some concrete tools. However, in the discussion, we also aim to connect the two subjects, to eventually conclude which actions are recommended to enhance students' disciplinary writing skill, regardless of subject-specificity.

CHAPTER 2

STUDENTS' DISCIPLINARY WRITING PROCESSES: A THINK-ALOUD STUDY*

Source-based writing is a common, but difficult task in history and philosophy. Students are usually taught how to write a good text in language classes. However, it is also important to address discipline-specificity in writing, a topic likely to be taught by content teachers. In order to design discipline-specific writing instruction, research needs to identify which reading and writing activities during the source-based writing process affect students' thought process quality and text quality, as assessed by content teachers. We conducted a think-aloud study with 15 (11th grade) students who performed two source-based writing assignments, each representative of its discipline. From the data, we derived 11 activities which we analyzed for duration, frequency, and time of occurrence. Results showed that the disciplines required different approaches to writing. For philosophy, the writing process was dominant and influenced quality, leading us to conclude that philosophical thinking and writing are intertwined. For history, the planning process appeared to be paramount, but it influenced text quality only, and not the quality of the thought process. In other words, historical thinking and writing appear to be separate processes. Our findings can be used to develop strategy instruction that reinforces better writing, adapted to discipline-specific writing processes.

* Chapter 2 is based on: Holdinga, C. C., Janssen, T. M., & Rijlaarsdam, G. C. W. (2021). The relationship between students' writing process, text quality, and thought process quality in 11th-grade history and philosophy assignments. *Written Communication, 38*(4), 544-586. <https://doi.org/10.1177/07410883211028853>

1. INTRODUCTION

Over the past two decades, technological development has created a new digital world full of text, which has forced us to substantially expand our literacy skills. Essentially, this has turned us all into writers. Research has shown that writing is now a critical success factor in nearly all domains, including school and professional environments (Freedman et al., 2016). In the Netherlands, the geographical context of our study, writing skills are a requisite in secondary education. This is particularly true in the social sciences because they tend to rely heavily on written assignments (Van Drie et al., 2016). However, most content teachers tend to focus on disciplinary thinking rather than writing (De Oliveira, 2011). Teachers in the disciplines central to our study – history and philosophy – are no exception. The Dutch history exam program does not include teaching students how to communicate a path of reasoning in its formal objectives. The philosophy exam program does include the acquisition of argumentative skills among its learning objectives, but it makes no specific mention of writing skills. In general, discipline-specific writing functions mainly as a means of assessment; students are expected to demonstrate their mastery of the discipline-specific thought process through a writing assignment. Writing in this sense is not a learning activity, but rather a window to students' thought processes.

Since writing is used so extensively to judge students' progress, one would expect an emphasis on teaching writing properly and extensively in secondary education. However, content teachers appear to pay little attention to writing instruction, writing process coaching or writing task design (Gillespie et al., 2014; Mottart et al., 2009). Content courses therefore seem to contribute little to the development of students' writing skills. This seems like a lost opportunity.

Paying more attention to writing skills in content courses might have other benefits as well. Research has shown that writing promotes learning (Graham et al., 2020; Graham & Perin, 2007; Klein, 1999; Newell & Winograd, 1995) and thus may strengthen, expand, and deepen students' content knowledge (Graham et al., 2020). Hence, writing can not only be used as a tool for assessment or as a goal in itself (to learn the mechanics of writing), but also as a learning activity, since the act of writing may support the acquisition of content knowledge. However, it is important to carefully consider what type of writing assignment is appropriate (Gillespie et al., 2014), since different writing assignments can have different learning effects (Applebee, 1984).

Furthermore, rhetoric and reasoning are not only genre-specific, but also discipline-specific (Bazerman, 1981; Klein & Boscolo, 2016).

Writing in school settings often involves reading (source-based writing), particularly in disciplines that use source materials, such as history (frequent use of historical sources) and philosophy (frequent use of philosophical source texts). This is why we have focused our study on these two disciplines. Source-based writing is a 'hybrid' task. Students are asked to produce a new text on the basis of reading one or more source texts (Spivey & King, 1989). To make such assignments useful for both content learning and improving writing skills, some kind of instruction is necessary. Simply adding general writing instructions is not sufficient, since addressing discipline specificity is crucial (Bazerman, 1981; Carter, 2007).

In order to be able to design such discipline-specific writing instructions, we first needed to explore the role of discipline specificity in source-based writing. We did this by exploring the reading and writing activities of 11th grade students performing source-based writing assignments for history and philosophy, and by asking the following questions: To what extent are these activities and their distribution across the whole process task-specific and therefore discipline-specific? Can the variation in approaches be related to variation in the quality of the resulting written text (the product) and to the quality of thought about a given issue (the path of disciplinary reasoning)? Our objective was to use the answers to these questions as an aid in designing instruction for 10th and 11th grade students in both disciplines.

1.1 Theoretical background

Previous research about reading and writing processes has given us valuable insights into the strategies used by expert readers and writers. Generally, expert readers are active readers, activating prior knowledge, predicting upcoming text content, drawing inferences, answering questions, reflecting on main points, constructing personal interpretations and images, self-explaining, and problem monitoring (Bråten & Strømsø, 2011; Chi, 2000; Pressley, 2002; Pressley & Afflerbach, 1995; Pressley & Harris, 2006). Furthermore, expert readers often go back and forth through the text, trying to meet their own standard of coherence (Van den Broek & Helder, 2017).

Expert writers typically plan and revise more than novices (Flower & Hayes, 1981; Hayes & Flower, 1986), as they tend to have more topic-, discourse- and language knowledge. They are also more active in monitoring their writing process (Ferrari et al., 1998). Moreover, experts make more

recursive and flexible use of reading and writing (Mateos et al., 2008; Lenski & Johns, 1997; McGinley, 1992). Furthermore, they try to transform their knowledge, unlike novices, who tend to restrict themselves to the repetition of knowledge (Bereiter & Scardamalia, 1987). With regard to experts' texts, research has shown that good writers write longer texts (Ferrari et al., 1998).

Whether these strategies are also crucial in discipline-specific writing is an unresolved question in need of further exploration. Previous research on reading in history does tell us, however, that a core competence is so-called 'sourcing' (Brante & Strømsø, 2018). Based on a think-aloud study with novice and expert historians who were asked to read particular source texts, Wineburg (1991) distinguished three main aspects of sourcing as carried out by experts: (1) checking for corroboration (is this plausible or likely considering other sources?), (2) checking credentials by sourcing (who wrote the text and when?), and (3) checking the contextualization (when and where did this happen?). Wineburg also noted that novices tend to regard texts as 'bearers of information' and hardly notice the source's features, whereas experts interpret texts in light of the source's characteristics and regard texts as social entities. Therefore, we can conclude that reading history texts requires more than general reading skills, as sourcing is considered a core skill in the discipline of history (Brante & Strømsø, 2018).

The core competency in philosophy is philosophical thinking, which entails (1) problematization (formulating questions and problems), (2) conceptualization (reflecting on philosophical concepts), and (3) argumentation (proposing and defending one's own arguments) (Tozzi, 2012). These skills can be practiced by writing, and particularly by writing argumentative texts (Corcelles & Castelló, 2017). Because a knowledge base of philosophical concepts is crucial to philosophical thinking, source texts play a role here as well. With regard to reading such sources, previous research suggests that expert philosophers are able to chunk the information from source material into categories (different views or bodies of opinion), with the help of metacognitive skills (Concepción, 2004).

In intervention studies, educational researchers have tried to equip content teachers with a knowledge base about writing, providing guidance not only on how to teach and coach writing (supporting the process of learning to write), but also on textual genres and other writing task conditions (supporting the process of writing to learn) (Klein & Boscolo, 2016). In the field of history, considerable research has been done on the question of how to encourage students to think and write as historians (De La Paz & Felton, 2010; Van

Drie et al., 2014; Wiley et al., 2014; McCarthy Young & Leinhardt, 1998). Likewise, in science, numerous studies have explored how to best stimulate awareness of scientific discourse and practice (Hand et al., 2002; Moje et al., 2004). Such research is based on the tenet that teachers should be aware of how disciplinary epistemologies define their expectations and understanding of what they regard as 'good writing'. Furthermore, teachers should be explicit about their own understanding of discipline-specific genre conventions if they are to improve students' writing abilities (Freedman et al., 2016).

1.2 Aim of the Present Study

Given the importance of source-based writing in history and philosophy, our aim was to explore reading and writing activities for discipline-specific source-based writing in both disciplines and to identify effective patterns in these activities. These insights will serve as a basis for the development of writing instruction, aimed at improving students' discipline-specific writing skills. Our main questions were as follows: Which patterns in source-based reading and writing can be related to (a) the quality of text produced by students, and (b) the quality of students' thought processes? Are there differences with regard to the previous question depending on the discipline under consideration (history versus philosophy)?

2. METHOD

2.1 Research Design

We set up a descriptive, within-subject, think-aloud study followed by reflective interviews, with students performing source-based writing in history and philosophy. We combined online (think aloud) with off-line (reflective interviews) data.

2.2 Participants

Fifteen Dutch students (11th grade, pre-university level, 10 females and 5 males, mean age: 16.8) from seven different high schools in the Netherlands participated voluntarily. We selected relatively high achievers, as we reasoned that the processes used by high achievers could serve as models for instructional design. Since experienced teachers are believed to make accurate judgments on students' achievement level (Südkamp et al., 2012), we asked teachers of history and philosophy to select these high achievers for our study. Without further guidance, the teachers relayed our request to those who they

regarded as 'good students'. Because we wanted to include students who were high achievers in both disciplines, we checked whether the students' grades were above average for both disciplines. Informed consent was obtained from all students.

2.3 Procedure

Data was collected by means of two think-aloud assignments (one for each discipline) presented in random order and with no time limit. Students were instructed to verbalize their thoughts and feelings while performing the tasks. Subsequently, we conducted follow-up reflective interviews. All sessions took place in convenient and quiet locations determined in consultation with the student. To test the feasibility of the procedure, we conducted a trial session with two students.

At the beginning of a session, the researcher explained the procedure to the student. During the think-aloud tasks, the researcher would visualize the student's thought process by arranging sticky notes on a piece of paper. On these sticky notes, the researcher described the student's activity (e.g., "reading" or "revising"). The students were told not to pay attention to this, because the process schemes would be discussed (to check accuracy and to clarify uncertainties) with the students afterwards, in a reflective interview. When necessary, the researcher (first author) prompted the thinking-out-loud process by posing questions (e.g., "What are you thinking now?"). All sessions took 40-50 minutes and were audio recorded.

2.4 Task Construction

In consultation with history and philosophy teachers, we selected two assignments from a textbook (philosophy) and an exam (history), ensuring both assignments featured typical characteristics of their respective discipline. Other selection criteria were: (a) inclusion of a source text, (b) similarity in genre, difficulty level, and time required, and (c) performability without prior content knowledge.

Task selection occurred in several rounds. After exploratory interviews with 21 teachers to identify disciplinary focus (unpublished data), we selected two assignments per discipline and discussed these with panels of two teachers. On this basis, we decided which of the two tasks to assign for each discipline. Finally, the two selected assignments were tested with two students.

Since our aim was to use typical disciplinary assignments, the assignments themselves were not entirely similar. Again, we strove for assignments that

reflected their respective discipline; the history assignment included a question about the usability and reliability of the presented source text (“is source text A pertinent to topic X?”), and the philosophy assignment contained a statement open for discussion (“to what extent is statement X true?”). The teacher panels confirmed that the two assignments were representative of their disciplines. In this article, we will therefore refer to the differences between assignments and between the reading/writing processes when performing these tasks as differences between the disciplines.

Because the key questions in the two assignments differed, the source texts were used differently as well. The history assignment required students to use sourcing skills, while in the philosophy assignment, which focused on students’ philosophical thinking skills, the source text was merely an aid to generate arguments. Although both assignments contained a single question requiring two arguments, in the history assignment those two arguments were interdependent. By contrast, the philosophy assignment could be divided into several parts to be dealt with independently. Lastly, the philosophy assignment was somewhat shorter (279 words) than the history assignment (427 words). See Appendix A for both assignments.

2.5 Data Preparation

2.5.1 Coding Scheme

We used both theory and research data to develop a coding scheme for analyzing the think-aloud protocols. The research we used to inform the coding scheme included research on general reading strategies (Merchie & Van Keer, 2014; Rogiers et al., 2019; Schellings et al., 2006; Vandeveldel et al., 2015); research on general writing strategies (Breetvelt et al., 1994; Kuhn et al., 2015), and on general metacognition (Nelson, 1996; Veenman, 2011); research combining reading and writing (Martinez et al., 2015; Mateos et al., 2008); research on disciplinary reading and/or writing (Brante & Strømsø, 2018), history reading and/or writing (Hof, 2015; Van Drie & Van Boxtel, 2008), philosophy reading and/or writing (Corcelles & Castelló, 2015); and research on general learning and reasoning (Bisra et al., 2018; Chi, 2000; Kuhn, 1991; Kuhn et al., 2015). We started from three main activities: reading, writing and metacognition, and then further subcategorized each of these activities into more detailed activities. Reading was subdivided into initial reading, rereading and analyzing the assignment or the source. Writing was subdivided into planning, writing, reading text produced so far and reviewing. Metacognition was subdivided into self-instruction, monitoring and evaluating. The subsequent

refinement of the coding scheme was an iterative process; we developed the coding scheme further during the first rounds of coding, resulting in additions and deletions of categories within the main activities. The final scheme is presented in Table 2.1.

2.5.2 Coding

All audio-recorded sessions were transcribed into written protocols by the first author of this paper. These protocols were subsequently divided into units in which a phrase or group of sentences was identified as a single thought process or approach representing a single activity. Every unit's duration was noted in seconds. The number of uttered words was between 537 and 2063 (history: $M = 1117$, $SD = 420$; philosophy: $M = 1067$, $SD = 453$). The number of units per protocol ranged from 19 to 132 (history: $M = 64$, $SD = 27$; philosophy: $M = 64$, $SD = 34$).

Interrater reliability was established by having a second independent coder code two randomly selected protocols, containing 53 units in total ($\kappa = .78$). An intra-rater reliability of .89 was established by re-coding a data subset (20%) one week after the coding of all protocols was completed.

Table 2.1. Coding Scheme

Code	Process activity	Category	Explanation	Example
M	Metacognition	Self-instruction Monitoring Evaluating	Telling yourself what to do next Monitoring part of the process Evaluating part of the process	<i>I am going to read the text first.</i> <i>Hmm, I still don't get what is expected of me.</i> <i>Is this a good example? Yes, I think so.</i>
R	Reading		Reading (parts of the) text	<i>((reads))</i>
RR	Rereading		Rereading (parts of the) text	<i>((rereads))</i>
TA	Task analysis	Task representation	Orienting yourself to the task or interpreting the assignment	<i>So, I need to give two examples.</i>
		Taking notes	Writing down keywords	<i>((takes notes))</i>
		Marking key words	Underlining or marking key words	<i>((marks key words))</i>
SA	Source analysis	Identifying source features	Noting author-specific phrasing and intentions and putting the source text in a context	<i>I see the text was written by a monk.</i>
		Drawing inferences	Making inferences or deductions based on the source introduction and source text.	<i>It says here that a kingdom was destroyed; that implies they were fighting.</i>
		Evaluating source content	Evaluating source text content, e.g. deciding what is important	<i>I think this part of the text is important.</i>
P	Planning	Generating	Generating ideas, reasons, arguments, counterarguments, rebuttals	<i>An example of altruism is giving up your seat to a senior citizen on the bus.</i>
		Organizing	Organizing content, connecting content elements	<i>So, this is an example that counters the statement.</i>
		Goal-setting	Setting clear goals	<i>Now I have to explain why the source is less suitable for the other chapter.</i>
PAU	Pausing		Silence of five seconds or longer	<i>(0.05)</i>

Table 2.1. Coding Scheme (Continued)

Code	Process activity	Category	Explanation	Example
W	Writing	Translating Transcribing Text	Mental text construction Text stated out loud Actual text written down	
RT	Reading text produced so far		Reading (parts of the) text produced so far	<i>((reads text produced so far))</i>
REV	Reviewing	Evaluating Revising	Evaluating translations or written text Changing translations or written text	<i>Wait, no, that's not the right word.</i> <i><-deleted word(s) or sentence></i>
O	Other	Researcher Interaction Affective responses Other	Researcher is speaking to the participant Interaction with the researcher Positive or negative responses from the participant Unclear	<i>Try to think out loud.</i> <i>Is it ok to use a marker?</i> <i>((laughs))</i> <i>Uh.</i>

2.5.3 Assessment of Text Quality

The quality of the students' texts was assessed using a holistic comparative method (Comproved online tool, <http://www.comproved.com>) by four panels of judges: (1) philosophy teachers ($N = 6$), (2) history teachers ($N = 6$), (3) L1 teachers assessing history texts ($N = 10$) and (4) L1 teachers assessing philosophy texts ($N = 5$). Since our goal was to determine relative quality, we opted for a comparative assessment method that would reliably result in a ranking of texts and an ordinal scale. Comparative judgment has proven to be a valid, reliable assessment method (Verhavert et al., 2019). Assessors made 30 random comparisons of texts, each time comparing two texts and deciding which text they thought was better, without specific criteria. All assessors were qualified secondary education teachers. We decided to include panels of L1 teachers on the assumption that these teachers would focus mainly on general text quality, whereas content teachers would focus more on content. The text quality assessment resulted in four rankings of the students' texts, one by each panel of judges (judges' reliability between .80 and .86). The rankings by content teachers and L1 teachers correlated (history and L1: $r = .78, p = .001$; philosophy and L1: $r = .60, p = .019$), which led us to conclude that the teachers all used similar quality criteria (per discipline) to assess the students' texts. We decided therefore to continue our analysis using only the content teachers' rankings, as we believe that in a study on discipline-specific writing, content teachers comprise the most valid rater group. We now know that content teachers partially agree with language teachers but that there is domain-specific variance.

2.5.4 Assessment of Thought Process Quality

Our second quality assessment focused on the students' thought process, based on the assumption that there might be a divergence between students' thought process and the text they produced. Thought process quality was assessed by two panels of judges: a panel of philosophy teachers ($N = 3$) and a panel of history teachers ($N = 3$). All assessors were qualified teachers in secondary education and/or teacher trainers. Each rater independently ranked the 15 protocols on thought process quality, disregarding the resulting text, holistically, without having been given any specific criteria ("Rank the protocols in order of the quality of the student's thought process"). We asked the judges to explain their rankings. Ranking correlations varied from .62 to .89 for pairs of history raters, and from .56 to .74 for pairs of philosophy raters.

For further calculations, we created a definitive ranking using median rank scores.

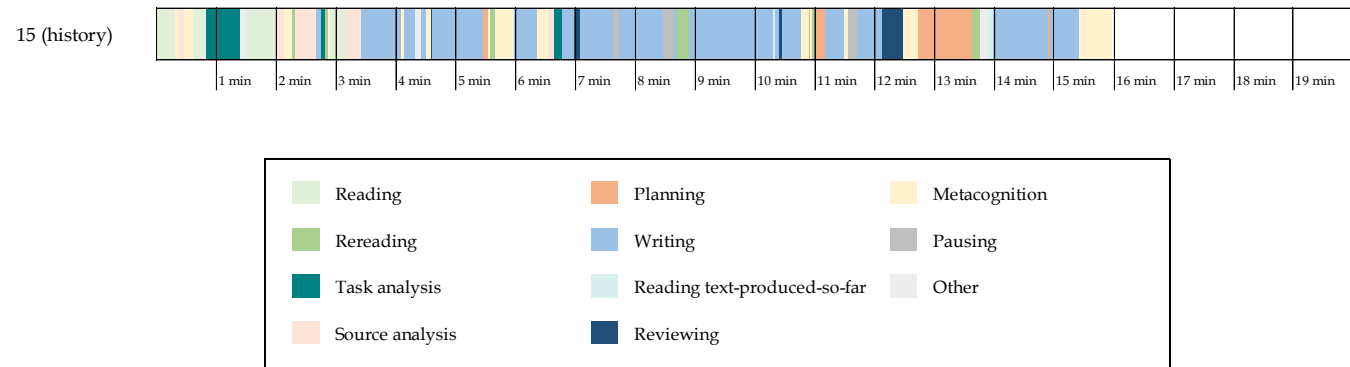
2.6 Data Analysis

To obtain an overall insight into each student's thought process, we analyzed the coded activities for duration and frequency, with duration referring to both absolute duration (in seconds) and relative duration (percentage of total time). Furthermore, we analyzed the number of words written (per writing spurt). As Figure 2.1 shows, we then visually represented the whole process (activities, duration) as process schemes, with colors representing the main activities from the coding scheme, and solid black lines indicating time units measuring one minute.

To refine our analysis, we subdivided the students' processes into three main segments, each covering approximately 33% of the total time spent on the process, and representing the beginning, middle and end of the process respectively. This allowed us to further explore at what point in the process a particular activity appeared to have an impact, since earlier research had shown that it is not only relevant what occurs, but also when in the process it occurs (Rijlaarsdam & Van den Bergh, 1996, 2006).

Furthermore, we calculated correlations between the process variables and the quality of the student's text and thought, both overall and for each main segment. These correlations served to spotlight aspects of the various processes that were sensitive to variation in text quality or thought process quality. When dealing with students who showed large discrepancies between text and thought process quality, we used these correlations as a starting point to conduct a qualitative analysis of their process and text.

Figure 2.1. Example of a Participant's Process Scheme



3. RESULTS

3.1 Correlations Between Process Activities and Quality

As Table 2.2 shows, we found significant correlations between process variables and text quality and thought process quality in both subjects, despite our limited number of participants. Process variables that appeared to influence quality included aspects of reading, source analysis, planning and pausing, writing, reading text produced so far, reviewing, metacognition, and total duration. We subsumed analyses of pausing under the planning category because pausing was associated with the idea generation process. This association was confirmed by significant correlations between pausing and planning variables (frequencies: $r = .46, p < .001$; percentage of time: $r = .30, p = .001$, duration: $r = .41, p < .001$).

Overall, it was clear that the students who spent more time thinking produced higher-quality texts (history: $r_s = .64, p = .010$; philosophy: $r_s = .54, p = .037$). As the students were not given a time limit, they took anywhere from 6 minutes 10 seconds to 20 min. 10 secs. (history: $M = 12.1, SD = 4.3$; philosophy: $M = 11.7, SD = 4.3$). We also noted different patterns for the two assignments. In the history assignment, we observed that 11 process variables were associated with text quality, three of which were also connected with thought process quality. None of the variables was uniquely connected with the thought process. Most of the variables that affected text quality were planning variables; students who planned more and spent more time on planning produced higher quality texts. In the philosophy assignment, text quality and thought process quality were more closely linked; eight variables had a bearing on text quality, four of which were also connected with thought process quality. Most of the variables affecting quality were writing variables.

Table 2.2. Overview of Correlations Between Writing Process Variables, Text Quality and Thought Process Quality

Variable	History				Philosophy				
	Text quality		Thought quality		Text quality		Thought quality		
	r_s	p	r_s	p	r_s	p	r_s	p	
Total	Duration	.64	.010**			.54	.037*		
M ^a	Frequency	.55	.033*	.53	.041*				
R	Duration					.52	.047*		
	% of time	-.68	.005**	-.55	.032*				
SA	% of time	-.53	.043*						
P	Duration	.64	.011*	.63	.012*				
	% of time					-.53	.044*		
	Frequency	.62	.013*						
PAU	Duration	.58	.024*						
	% of time	.55	.035*						
	Frequency	.57	.027*						
W	Duration	.55	.035*			.89	.000**	.63	.012*
	% of time					.54	.038*		
	Frequency					.56	.029*	.63	.012*
	Number of words	.67	.007**			.92	.000**	.61	.016*
RT	Duration							.57	.025*
	Frequency					.62	.014*	.67	.007**
REV	Duration							.52	.048*

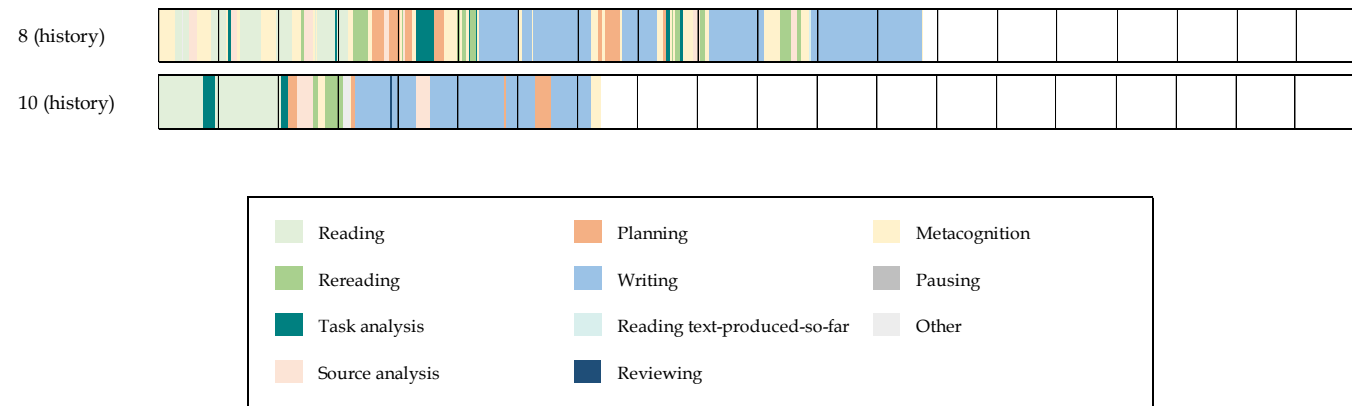
Note. M = Metacognition, R = Reading, SA = Source Analysis, P = Planning, PAU = Pausing, W = Writing, RT = Reading Text produced so far, REV = Reviewing.

^aCategory: Evaluating.

3.2 Metacognition

Metacognitive activities were generally brief ($M = 4.3$ s, $SD = 2.1$ s). The frequency of metacognitive segments varied markedly from participant to participant. For example, Student 8 expressed 35 metacognitive thoughts while carrying out the history assignment, whereas Student 10 expressed such thoughts only three times (Figure 2.2).

Figure 2.2. Comparison of Two Students' Use of Metacognition During the History Assignment



Metacognition frequency correlated with the total number of segments ($r = .46, p = .011$), which indicates that students whose protocols took longer also engaged in more metacognitive activities. As Table 2.2 shows, in the history assignment, evaluation frequency correlated with both text quality ($r_s = .55, p = .033$) and thought process quality ($r_s = .53, p = .041$). This indicates that students who evaluated more frequently during the process, produced higher quality texts and had higher quality thought processes.

Table 2.3 depicts the correlations between aspects of metacognition and quality in the three main segments. The Table illustrates that students who produced higher quality texts and thought processes evaluated more often, particularly during Segments 1 and 2. For philosophy, metacognition frequency only correlated with thought process quality during Segment 1 ($r_s = .61, p = .016$).

Metacognitive activities mostly served as a catalyst for switching to another activity (80%) rather than an interruption of one and the same activity (20%). In the history assignment, metacognitive activities tended to prompt the student to switch to rereading. This was the case in all three of the main segments. However, each segment featured a different secondary focus: during Segment 1, the students' secondary focus was on task analysis; during Segment 2 on writing; and during Segment 3 on reading the text they had produced so far, and on reviewing and pausing.

In the philosophy assignment, metacognitive activities prompted fewer switches to rereading the assignment and source text, and more switches to writing, irrespective of the segment. Our analysis of the separate segments of the philosophy assignment yielded a pattern similar to the pattern we found in history, but with different secondary foci: during Segment 1, the students' secondary focus was on rereading and task analysis; during Segment 2 on planning; and during Segment 3, on reading what they wrote so far. These differences in main and secondary foci between the two disciplines indicate that the assignments may trigger different approaches.

3.3 Reading

The concept of 'reading' contained various kinds of reading activities: reading, rereading, task analysis and source analysis. The only correlations for quality we found were with reading and source analysis (see Table 2.2). Table 2.4 presents the correlations between reading variables and quality, analyzed for each of the three main segments.

Table 2.3. Correlations Between Metacognitive Activities, Text Quality and Thought Process Quality

Variable		History								Philosophy							
		Text quality				Thought quality				Text quality				Thought quality			
		0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
Monitoring	Frequency		+					+									
Evaluating	Frequency	+	+	+				+	+								+

Note. 0 = total process, 1 = Segment 1, 2 = Segment 2, 3 = Segment 3.

Table 2.4. Correlations Between Reading Activities, Text Quality and Thought Process Quality

Variable		History								Philosophy							
		Text quality				Thought quality				Text quality				Thought quality			
		0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
Reading	Duration									+	+						
	% of time	-	-			-	-										
Source analysis	% of time	-		-													

Note. 0 = total process, 1 = Segment 1, 2 = Segment 2, 3 = Segment 3.

3.3.1 Initial Reading

In terms of absolute reading time, there was a significant difference ($t = 6.82$, $p < .001$) between the two disciplines (history: $M = 106.47$ s, $SD = 15.55$ s; philosophy: $M = 80.93$ s, $SD = 6.79$ s). This difference can be explained by the length of the respective assignments: at 427 words, the history assignment was longer than the philosophy assignment (279 words). When comparing relative duration, the difference between the disciplines disappeared (history: $M = 16.6\%$, $SD = 6.6\%$; philosophy: $M = 13.0\%$, $SD = 4.4\%$; $t = 1.97$, $p = .069$). For history, the percentage of time spent on reading had a negative effect on text quality ($r_s = -.68$, $p = .005$) and thought process quality ($r_s = -.55$, $p = .032$). By contrast, in philosophy reading process duration appeared to be a quality indicator, that is, students who spent more time reading produced higher quality texts ($r_s = .52$, $p = .047$). Obviously, this was observed only during Segment 1 (Table 2.4), because initial reading only takes place at the beginning of the task.

The students dealt with the initial reading phase in various ways. Some approached this part of the task very straightforwardly, reading without pause or interruption and seemingly no goal orientation. After this first readthrough, some students struggled to understand the text and had to

reread it, but others seemed to grasp the ideas expressed in the text in a single reading and were able to start generating a response immediately afterwards. For example, Student 6 (Figure 2.3) read the text without any interruption, and then immediately started the planning phase. Other students took a more complex approach and alternated reading with rereading, task analysis or source analysis. This approach might point to a desire to fully understand the text, for example in the case of participants who immediately wanted to interpret and analyze what they read, as Student 15 did. Alternatively, a more complex approach might also point to reading problems, with participants failing to understand or misinterpreting parts of the text, as seemed to be the case for Student 8. However, we did not find a relationship between reading approach and text quality or thought process quality.

3.3.2 Rereading

Some students reread the assignment and the source text more than once and spent a relatively long time on this, while others did not reread at all, as the process schemes in Figure 2.4 illustrate. For example, Student 11 reread the text several times in both disciplines, whereas student 13 did not reread in either of the disciplines. Although these examples might indicate 'individual preference', other students took a different approach depending on the discipline. For example, Student 7 adjusted their rereading behavior to the discipline (or the assignment).

When we analyzed the rereading activities for the three segments of the process, we found that both duration and percentage of time spent were stable across all three segments in the history task, but were not stable in the philosophy task. In philosophy, students spent a significantly ($t = 3.34, p = .005$) lower percentage of the time rereading ($M = 0.91\%$; $SD = 1.06\%$) than in history ($M = 2.93\%$; $SD = 0.63\%$). However, in neither discipline (neither overall, nor in a specific segment) was the time they spent rereading connected to the quality of the text they produced or to the quality of their thought process.

We also noted that some students took more time to reread than to do a first readthrough. These students often did this for both assignments, which we took to indicate a 'personal strategy'. This assumption was supported by the correlation between the percentage of time spent rereading for history and for philosophy ($r = .51, p = .050$), which occurred mainly during Segment 3 ($r = .61, p = .016$).

Figure 2.3. Different Approaches to Initial Reading

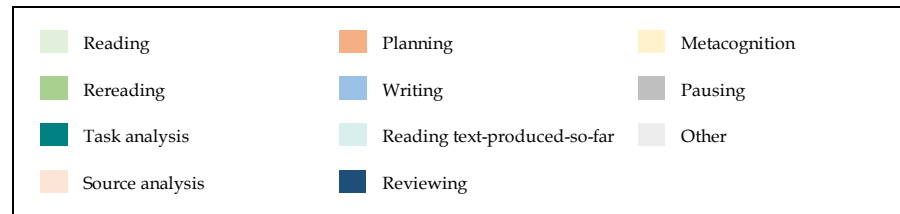
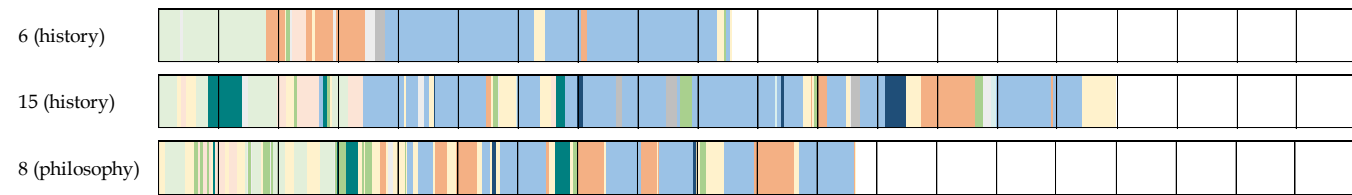
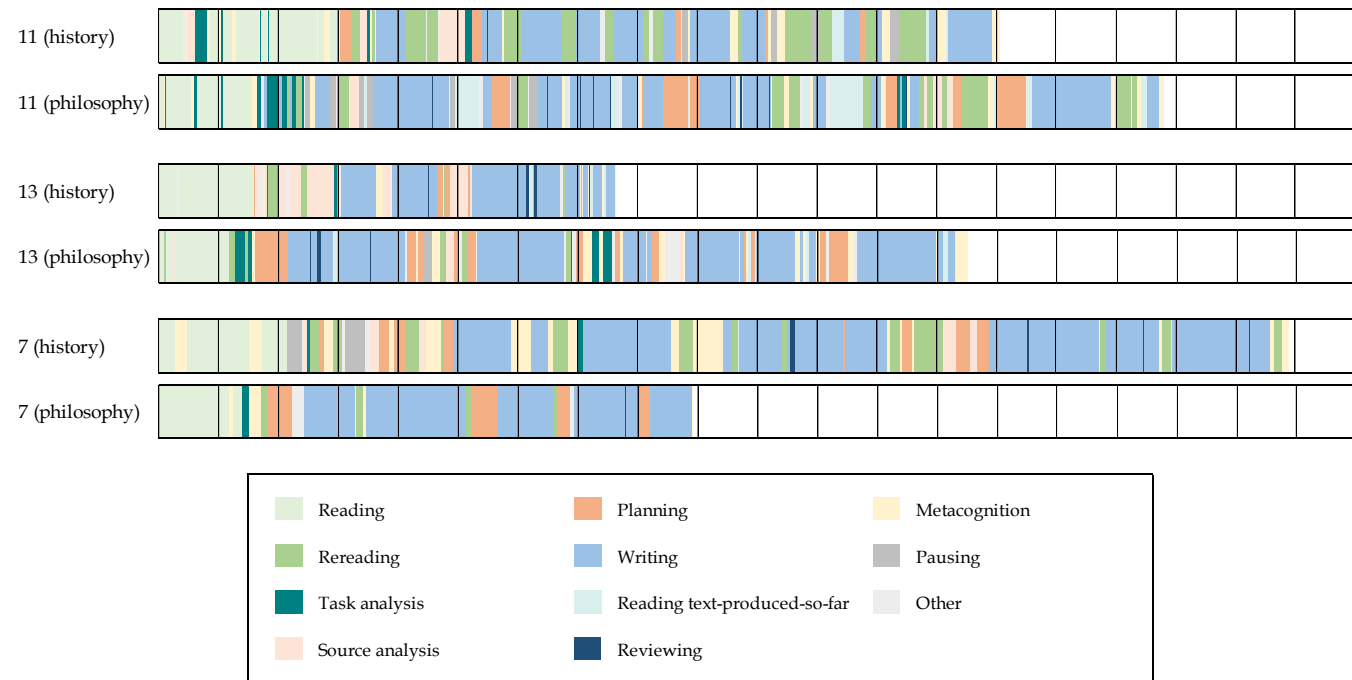


Figure 2.4. Differences in Rereading Behavior



3.3.3 Source Analysis

Some students analyzed the source text extensively, while others did not. Overall, students spent a significantly larger percentage of the time ($t = 7.94$, $p < .001$) analyzing the history source text ($M = 7.4\%$, $SD = 4.27\%$) than the philosophy source text ($M = 1.58\%$, $SD = 1.87\%$). This difference was observable in Segments 1 and 2.

The percentage of time spent on source analysis negatively correlated with text quality for history ($r_s = -.53$, $p = .043$). In other words, students who performed better on text quality seemed to have spent less time reading and analyzing the source. This applied only to Segment 2 (see Table 2.4).

3.4 Writing

For our research we defined the concept of 'writing' to include planning, pausing, writing, reading text produced so far and reviewing. Table 2.5 shows the correlations between writing variables and quality in the three main segments.

Table 2.5. Correlations Between Writing Activities, Text Quality and Thought Process Quality

Variable	History						Philosophy						
	Text quality			Thought quality			Text quality			Thought quality			
	0	1	2	3	0	1	2	3	0	1	2	3	
Planning	Duration % of time	+			+							+	
	Frequency	+							-			+	
Pausing	Duration % of time	+	+										
	Frequency	+	+				+						
Writing	Duration % of time	+		+					+	+	+	+	+
	Frequency	+							+	+	+	+	+
	No. of words	+			+				+	+	+	+	+
	Duration % of time								+	+		+	+
Reading text produced	Duration % of time								+	+		+	+
	Frequency								+	+		+	+
Reviewing	Duration % of time					+		+				+	+
	Frequency								+				

Note. 0 = total process, 1 = Segment 1, 2 = Segment 2, 3 = Segment 3.

3.4.1 Planning and Pausing Time

We analyzed the planning and pausing phases for duration, percentage of total time spent, and frequency. We saw significant differences between the

disciplines ($t = -2.35, p = .034$) in terms of the percentage of time spent on planning (history: $M = 9.27, SD = 4.77$; philosophy: $M = 14.35, SD = 7.68$). Students spent relatively more time planning for philosophy than for history. For philosophy, the relative time spent on planning correlated negatively with text quality ($r_s = -.53, p = .044$). Another significant difference ($t = -2.28, p = .039$) between the two disciplines pertains to how frequently the students paused (history: $M = 1.80, SD = 1.86$; philosophy: $M = 4.07, SD = 4.65$).

Most of the planning and pausing aspects seemed to influence text quality in history, but not in philosophy. In the history assignment, the students who planned more often and for longer periods of time, and who paused more frequently and for longer periods of time, produced better texts. The duration of planning also correlated positively with the quality of their thought process ($r_s = .63, p = .012$).

In our analysis of the interaction between generating ideas and writing, we noticed that some of the students seemed to generate while writing, instead of writing down what they had first generated. Student 14 remarked upon this, saying: "This pops up while I'm writing." As a consequence, their process included very few idea-generating sections (Figure 2.5). It seems the writing process fostered their thought process. Student 15 also displayed no obvious idea-generating process starting the writing process (see Figure 5).

There was no connection between this strategy and quality, although we did find a negative correlation between percentage of time spent on planning and text quality for philosophy ($r_s = -.53, p = .044$). Students who spent relatively less time on planning wrote better texts. This was not connected to a specific segment.

3.4.2 Writing Time

The time the students spent on the writing process varied widely in both disciplines (history: $M = 297.5$ s, $SD = 144.1$ s; philosophy: $M = 287.1$ s, $SD = 109.5$ s) and correlated strongly with the length of the text they produced (history: $r = .88, p < .001$; philosophy: $r = .94, p < .001$). In other words, the students who spent more time writing also wrote longer texts. We also observed that the students who produced higher quality texts seemed to spend more time writing (history: $r_s = .55, p = .007$; philosophy: $r_s = .89, p < .001$) and produced longer texts (history: $r_s = .67, p = .007$; philosophy: $r_s = .92, p < .001$). Text length ($r_s = .61, p = .016$) and writing time ($r_s = .63, p = .012$) also correlated with thought process quality for philosophy, but not for history.

In 16 out of 30 cases, the students started writing as early as Segment 1, although in six of the cases this only involved writing 'starter sentences' (e.g., rephrasing the question). Conversely, in 10 out of 16 cases this early writing contained actual content (history: three students; philosophy: seven students). For philosophy, students who spent a larger percentage of the time writing in Segment 1 scored better on text quality ($r_s = .72, N = 15, p = .003$) and thought process quality ($r_s = .59, N = 15, p = .022$).

3.4.3 Writing spurts

In our analysis of the students' writing in terms of frequency of spurts and number of words per spurt, we found that most students wrote in spurts of 13-15 words on average (history: $M = 12.8, SD = 5.5$; philosophy: $M = 15.4, SD = 6.4$). Some students managed to write the full text in only a few (longer) spurts of writing. This happened in both disciplines. It seems that these students had a personal approach to writing. For philosophy, the frequency of writing spurts correlated with text quality ($r_s = .56, p = .029$) and thought process quality ($r_s = .63, p = .012$), meaning that the students who wrote in more spurts, produced higher quality texts and had higher quality thought processes. This was visible in Segments 1 and 3. These findings can be substantiated by the negative correlation between the number of writing spurts and mean number of words per spurt ($r = -.60, p < .001$), i.e., the more spurts, the shorter their length. This pattern of numerous short spurts of writing might be an indicator of recursivity.

3.4.4 Reading Text Produced So Far

Some students read the text they produced so far quite often and some did not reread what they wrote at all. This finding was true in both disciplines. Reading text produced so far could be categorized as a strategy to either (1) generate new content or continue writing (76% of the cases) or (2) review or evaluate what had been written so far (24% of the cases). Students who read what they wrote so far early on in the process (in Segments 1 and 2) mainly did this as a strategy to create content. For philosophy, the frequency of reading text produced so far correlated to text quality ($r_s = .62, p = .014$) and to thought process quality ($r_s = .67, p = .007$). For history, there was no overall correlation between reading text produced so far and quality. However, in Segment 2, there was a correlation between duration, percentage of total time spent, and frequency of reading text produced so far with thought process quality (see Table 2.5).

Figure 2.5. Process Schemes of Students Who Generated Ideas While Writing



3.4.5 Reviewing

Some students did not review (revise/edit) their text at all, while others reviewed it quite often. This occurred in both disciplines. Reviewing occurred both during the writing process (in Segment 1 or 2) (history: 46%, philosophy: 54%), and at the end of the process (in Segment 3) (history: 56%, philosophy: 44%). Overall, the reviewing process was connected to thought process quality in philosophy but not in history (see Table 2.2). The correlation between the duration of the reviewing process and thought process quality ($r_s = .52, p = .048$) might indicate that the students' thought process was aided by the reviewing process. In the history task, reviewing was connected to thought process quality in Segment 1 (see Table 2.5). In the philosophy task, reviewing was also connected to text quality, but again, only in Segment 1. The latter two findings were linked to early writing, as students who only started writing in Segment 2 obviously would have had nothing to review in Segment 1.

3.5 Indicators of Good Writing

Since text quality and thought process quality correlated for both disciplines (history: $r_s = .81, p < .001$; philosophy: $r_s = .61, p = .015$), students with higher quality thought processes generally produced higher quality texts. Figure 2.6 and Figure 2.7 show the data points for both correlations, with 15 being the highest and 1 the lowest possible quality rank score. Next, we marked the outliers, i.e., the students who 'underachieved' in the sense that they showed high-quality thinking, yet produced a relatively weak text. We compared their process schemes with those of students who performed better on the writing task (Figure 2.8). The underachievers' performance will be discussed in more detail in Sections 3.5.1 (History) and 3.5.2 (Philosophy).

Figure 2.6. Relationship Between Text Quality and Thought Process Quality for History

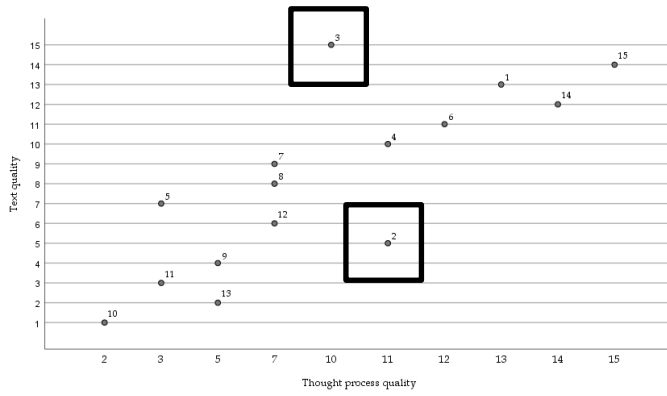


Figure 2.7. Relationship Between Text Quality and Thought Process Quality for Philosophy

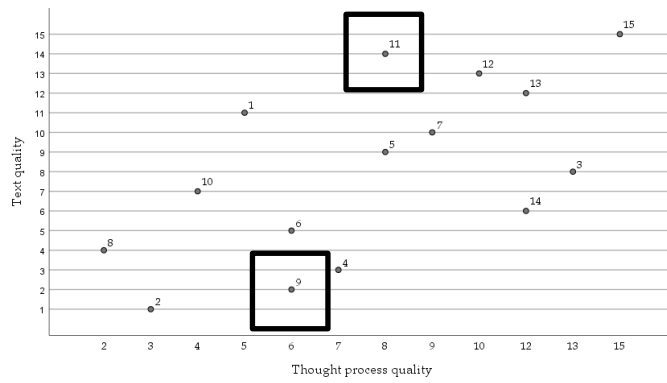
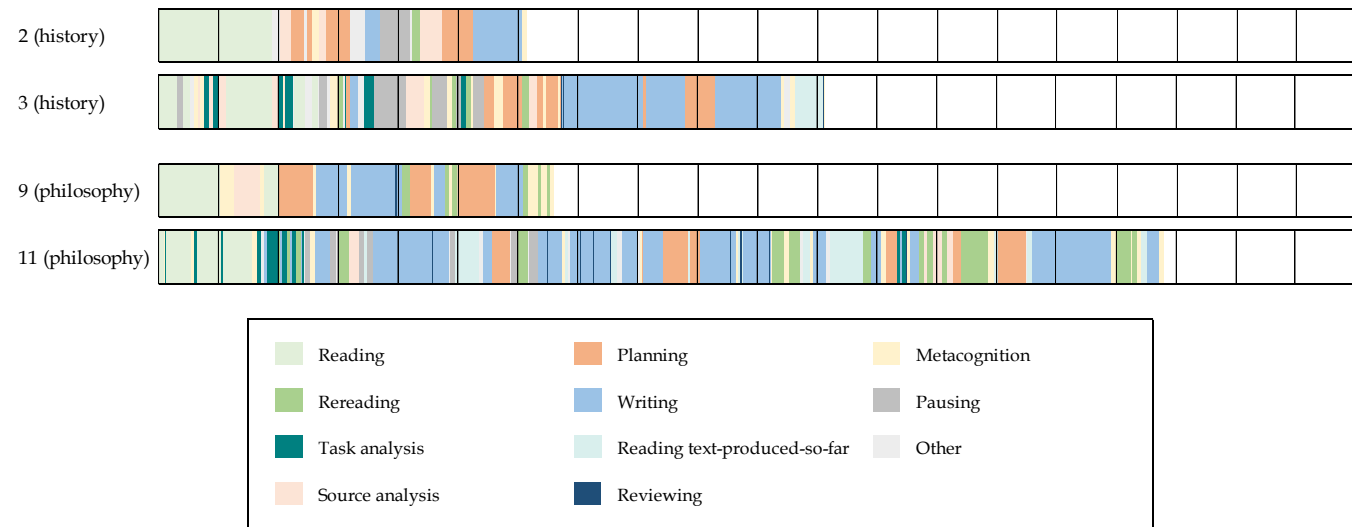


Figure 2.8. Process Scheme Comparison: Underachievers (2, 9) vs. Good Writers (3, 11)



3.5.1 History

Student 2 wrote a relatively weak text (scored 5 out of 15), although the quality of their thought process was considered relatively high (scored 11 out of 15). After a reading phase without interruptions, this student analyzed the source text, alternating with generating ideas. Next, the student wrote down the phrase "Chapter 1", then paused, thought about and analyzed the source text more extensively, and eventually wrote down their complete answer in a single 21-word spurt of writing. After that, the student said: "I think that's it". Their process showed decent procedural knowledge of source analysis as acquired in history class. For example, immediately after reading the assignment they started identifying source features:

Student 2 *First, I would look at who wrote the source text. I see that the text was written by a monk, so the second part of the question applies here.*

However, the student's text referred to this source analysis very succinctly, which resulted in a relatively short text that was not representative of the arguments the student developed in their thought process:

Student 2's text *Chapter 1: less usable, written down almost 100 years after the event, biased.
Chapter 2: usable because the source text was written by a monk.*

Student 3's thought process quality (scored 10 out of 15) was similar to that of Student 2, but they wrote the best text in this sample. As Figure 2.8 shows, this student's process was more complex. Their reading phase was interrupted by note taking and source analysis, which continued during the second segment. According to one of the assessors of thought process quality, the thought process score suffered from the student's relatively drawn out reading process. The student struggled to make sense of the assignment and the source text. Consider the following excerpt, recorded five minutes into the process:

Student 3 *Um, honestly, I still don't understand the assignment
(rereads) So, for which chapter is the source more useful.
(0.05)
Ah, so I need to choose, yes, I have to make a choice.*

Furthermore, Student 3 did not use explicit procedural knowledge in their process, for example, they did not explicitly mention any of the source text's features. However, when this student appeared to start feeling confident

about their reading and reached their ‘standard of coherence’ (Van den Broek & Helder, 2017), they managed to generate and formulate an answer relatively fast in five spurts of writing, interspersed with revising errors. At the end of the process, they reread their entire text and revised only a minor error. The resulting text was correct, concise and well-structured.

Student 3’s text *The source is useful for the chapter on a monk’s world view, because he does not objectively relate the course of the war, he says "Oswiu appealed to God’s grace and help" and it turns out that his army wins. So, while this information it is not useful to determine the actual course of the war, it is useful for representing the monk’s view of the world, namely that you have to believe in God and if you do not, you will lose, no matter how strong your army is.*

3.5.2 Philosophy

In the philosophy task, Student 9 showed average thought process quality (scored 6 out of 15) and produced one of the weakest texts (scored 2 out of 15). Student 9 analyzed the source text during the reading phase, immediately after reading it. Subsequently, they displayed a three-step, non-recursive process, generating an answer to the first part of the question and phrasing this in three spurts of writing. The student tried to generate two separate arguments and wrote these down in one spurt of writing each. Lastly, they reread the assignment and checked the text for completeness. For this student, the discrepancy between thought quality and text quality appeared to be caused by the transformation of their thought process into writing. Consider the following example:

Student 9 *((reads)) An example that supports the statement that people are selfish. An example is people who donate 100 euro to a charity for orphans, although they don’t have any connection to those children. They think they do something good, but nine times out of ten, people are only concerned with making themselves feel better. So, that’s what I will write down.
((transcribes)) Example: people who donate money.
((writes)) e.g., people who donate money.*

Student 9’s text ended up being short and lacking any of the good explanations they gave while thinking out loud:

Student 9’s text *People are not always selfish. I think that many actions are selfish, but if an individual has ties with another ‘individual’, it is not just for themselves. e.g., people who donate money.
example: parents who take care of their children are not selfish.*

By contrast, Student 11, who also displayed an average thought process (scored 8 out of 15), produced a much higher quality text (scored 11 out of 15). During the reading phase, this student marked key words and analyzed the assignment by taking notes. Student 11 switched between activities often, sighed from time to time, and reread the source text over and over, which we took to mean they found the assignment difficult. This student started writing early (in Segment 1) and the writing process itself was fragmented (18 spurts). The result was a well-structured and quite lengthy text. At the end of the process, Student 11 evaluated their text based mainly on content.

Student 11's text *People are just selfish, because when you leave money to someone, for example, you only leave this to your children, because, out of self-interest, you don't want this to go to waste (as in: your hard work has been for nothing / your money goes to someone else). By giving your money to your children, you protect your own reputation (this is your only way of maintaining control over this after your death). So, this is pure self-interest.*
People are not just selfish, because sometimes people work hard especially for their children. They don't leave their money to their children to protect their own reputation, but to give their children opportunities and so indirectly pass on your genes.
People are not just selfish, because we are social animals. We need each other to survive and we are willing to help each other. This is evident, for example, from the fact that we have children and are willing to put money and effort into someone else's life, as described in the source text.

4. DISCUSSION AND CONCLUSIONS

The present study explores which reading and writing activities during source-based writing about history and philosophy affect (text and thought process) quality as assessed by content teachers. We approached this topic by conducting a think-aloud study with 15 participants (11th grade students) who performed a source-based writing assignment in both disciplines. From all the data gathered, we derived 11 activities, which were subsequently analyzed for relative and absolute duration as well as for frequency of occurrence and time of occurrence (beginning, middle or end of the process, i.e., during Segment 1, 2 or 3).

The results show that students approached the two assignments somewhat differently in terms of the three main activities: metacognition, reading and writing. For history, metacognition often acted as a bridge to rereading the assignment or source text, whereas for philosophy, metacognition more often segued into writing. In terms of reading, students spent more time rereading (in Segment 2) and more time analyzing the source text when doing the history assignment than when working on the philosophy assignment. With

regard to writing, the students seemed to tackle the history and philosophy assignments in similar ways, although they spent more time planning and pausing more frequently when working on the philosophy assignment.

Furthermore, the students' activity patterns and these patterns' relationship to quality differed for the two disciplines. A possible explanation for this could be that the two assignments require different dispositions of the writer towards the topic (Galbraith & Baaijen, 2018). For philosophy, the writing process was dominant and influenced both text and thought process quality, indicating interwovenness. Yet for the history assignment, the planning process seemed to be most important, but influenced only text quality. Despite a correlation between the two quality measures as a whole, the quality of text and thought process appeared unconnected. Thought process quality (and its correlation with text quality) thus depends on other factors than the ones we included in our analysis.

Overall, the metacognition and reading processes seemed to have only marginal impact on quality, even though other research has shown that expert readers tend to have a more active, recursive reading style (Pressley & Harris, 2006). However, our study did not support such a correlation between the students' reading approach and the quality of their writing or thought process. However, this might be due to the relatively short reading phase. Another possible explanation for our findings could be that all the students eventually managed to reach their 'standard of coherence' (Van den Broek & Helder, 2007), since it would have been almost impossible to write down an answer without reaching a level of understanding first. Monitoring and evaluating were influential in the beginning and the middle of the process (i.e., in Segments 1 and 2), when reading takes place. This indicates that an active reading style might be beneficial anyway. With regard to rereading, we saw that the need to reread the source text was greater in history than in philosophy. This was probably due to the nature of the assignment, since source analysis was a prerequisite for history, whereas it was not for philosophy. However, the rereading behavior in history could also be interpreted as an indicator of recursivity in the process. Some students spent more time rereading than they spent reading initially; this could suggest that those students were using rereading as a coping strategy to generate ideas, for instance, when their thinking or writing process stagnated (Bereiter & Scardamalia, 1987).

Regarding planning, it is known that writers can use a variety of methods to retrieve and generate ideas. These methods include the generation of ideas prompted by reading the assignment, the activation of new ideas when ideas

are translated into text, or the development of ideas while structuring or revising the text (Van den Bergh & Rijlaarsdam, 1999). Looking at the negative association between time spent on planning and text quality for philosophy, we therefore recommend that students working on a philosophy assignment try to generate ideas while writing. However, for students carrying out a history assignment, we recommend planning carefully and writing down answers after a thorough planning phase in order to reinforce a recursive pattern of planning and writing.

Overall, the writing process seems to be crucial for philosophy. Students who wrote relatively high-quality texts also produced longer texts, spent more time writing, and reviewed more frequently, which are all characteristics of expert writers (Ferrari et al., 1998). Moreover, the writing process seemed to enhance the thought process; these processes were closely intertwined. This is corroborated by the finding that starting the writing process during the first segment was associated with better texts: a finding that suggests writing and thinking co-develop.

To explore indicators for good writing, we also conducted a qualitative analysis of 'underachievers', i.e., students whose thinking and reasoning was sound, but who wrote relatively weak texts. Our analyses showed that these students tended to write very succinctly and that their texts lacked proper linking phrases and essential explanations that these students had voiced in their thought process.

Our conclusions must be viewed in light of the limitations of the present analysis. In the explanation we gave of our methodology, we discussed our reasons for presenting the differences between the assignments as differences between the two disciplines. We carefully selected assignments that were representative of their respective disciplines, implying that the differences in the assignments were indicative of actual disciplinary differences. However, since we used only one assignment per discipline, our results might also be attributable to the characteristics of these assignments. We are aware that this limits the validity of study, but we still believe that our results are very useful because our research design explored students' intra-individual differences. These differences can be used to inform instructional design, which was our end goal.

A second limitation has to do with the scale of our study. Our small-scale design (i.e., 15 students) with high-achieving students for both disciplines allowed us to make an in-depth analysis of activities, but prevented us from drawing more general conclusions. Also, the think-aloud method has side-

effects that can play a role when researching metacognitive activities. Despite our precautions, students might have felt unsafe, and this in turn might have affected the data we collected and thus induced a possible bias. This is one of the well-known limitations of the think-aloud method (e.g. Ericsson & Simon, 1980).

Still, we believe our results are valuable, as our study corroborates the conclusions of other research regarding process characteristics in source-based writing (e.g. Mateos et al., 2008). Moreover, by including two different disciplines, we were also able to obtain an understanding of discipline-specific writing. And finally, our reflexivity as researchers (Berger, 2015) also played a role. The research team do not consider themselves true experts in the field of history or philosophy, because these fields were not the main scope of our study. However, our position as writing experts provided us with an opportunity to analyze discipline-specific writing for either discipline from a more distant point of view.

As far as classroom practice is concerned, the differences between history and philosophy we identified in this study shows the importance of addressing discipline-specificity in writing tasks in secondary education. Philosophical thinking and writing are intertwined and parallel processes, which suggests that writing is an appropriate tool for learning in philosophy, and discipline-specific writing instruction might foster writing to learn. However, historical thinking and writing turned out to be separate and partly serial processes. The students' texts often did not include their historical reasoning; the texts merely reflected the outcome of the students' reasoning. Hence, it seems valid to ask whether writing instruction in history is likely to contribute to students achieving our teaching goals. We therefore recommend further exploration of the relationship between what we aim to achieve with discipline-specific writing assignments and what we actually accomplish.

CHAPTER 3

WRITING TO LEARN HISTORY: AN INSTRUCTIONAL DESIGN STUDY*

This study reports on the design and evaluation of an instructional unit, aimed at improving secondary school students' disciplinary writing in history. Central to this design was the replacement of conventional workbook exercises by evaluative source-based writing tasks which were co-developed with participating history teachers. Additionally, an instructional unit to teach students a discipline-specific Read-Think-Write strategy based on previous research was designed. Two history teachers implemented the evaluative tasks and the strategy instruction in their 11th grade history classrooms in a trial intervention study with a switching panels design. Pre, mid, and post-testing consisted of the teacher-designed evaluative writing tasks (ca. 200-300 words), which were analyzed on holistic quality, content quality, quality of structure, and text length. Results showed effects in the second panel for content quality.

In this paper we elaborate on the design of this strategy and the instructional design, as well as the design principles underpinning these. Based on the trial study, we present recommendations for redesign in order to optimize practicality and effectiveness of the instructional unit.

* Chapter 3 is based on: Holdinga, C. C., Van Drie, J. P., Janssen, T. M., & Rijlaarsdam, G. C. W. (2023). Writing to learn history: An instructional design study. *L1-Educational Studies in Language and Literature*, 23(1), 1-44. <https://doi.org/10.21248/11esll.2023.23.1.526>

1. INTRODUCTION

Subject areas in upper secondary education demand a high level of literacy. Students are presumed to be able to read, and write, complex texts which are tuned to the relevant discipline. Improving literacy should thus play an important part in education; this insight has been shared by educational researchers for many years now. Around 1980, a general approach to teaching literacy was common: it emphasized the value of literacy instruction in content area classes, to help improve literacy. In this view, subject area teachers should adopt general reading and writing strategies from the language classroom in their own subjects.

However, this approach has proven to be insufficient (O'Brien, et al., 1995). Over the past decades, educational researchers such as Moje (2008) have argued that secondary content literacy should focus on subject areas, instead of on general literacy. They called for a change: content should always be put first, since learning in a subject area entails the understanding of the norms of practice for producing and communicating knowledge in that specific discipline. Disciplines can thus be considered discourse communities which students must navigate. This calls for disciplinary strategies, which should be emanated from content.

In the Netherlands, it is not a shared opinion among teachers that literacy should be part of every subject area, as an intrinsic element of the subject curriculum. Literacy development is still mainly considered to be the responsibility of the language department. History teachers share this latter view, even though it is well-known that history is a subject which demands extensive reading and writing (Mottart et al., 2009).

In the current chapter, we report on the design and evaluation process of a lesson series aimed at secondary school (11th grade) students' historical writing, with one focus in particular: learning remains paramount. With this study, we aimed to contribute to educational research by providing insights in effective elements of intervention research, by describing design principles and learning activities explicitly, and thoroughly, following a constructive trend initiated by other researchers in the educational domain (e.g. Schrijvers et al., 2019; Van Ockenburg et al., 2021a).

1.1 Relevance of Teaching Disciplinary Literacy

When literacy is overlooked in the content classroom, disciplinary knowledge of writing remains out of sight, although this knowledge of discipline-specificity is crucial (Carter, 2007). For example, history students need to

contextualize a phenomenon in their text, use sources to build an argument, and discuss reliability of sources. It is reasonable to expect such a 'historical lens' to be a history teacher's responsibility, since it is inherent to the domain of history.

However, content area teachers hardly spend time supporting students' writing process, nor do they provide writing process instruction (De Oliveira, 2011; Gillespie et al., 2014; Mottart et al., 2009). This may be logical from the content teacher's point of view, since it is not a formal objective of the Dutch history program to teach students how to communicate history (CvTE, 2022). Teachers tend to align their assignments with the official curriculum and what is asked in the final exam: in the Netherlands, that is, answering short open-ended questions. These questions generally focus on assessment of historical content knowledge and specific history skills, such as use of sources, or on metacognitive concepts, such as changes or causes. As a result, assignments comparable to the Document Based Question (DBQ), which is an exam task in the US that highly addresses students' disciplinary writing skills, are much less common in the Netherlands.

Yet, also in other countries, teachers struggle with discipline-specific literacy instruction (Nokes, 2010; Ragland, 2007). One explanation might be that teachers feel uncertain about providing reading-writing instruction, since it has had no prominent place in their teacher training either; literacy support seems underestimated and underrated (O'Brien, et al., 1995).

Meanwhile, teachers generally do require their students to write elaborated and well-structured texts. Students who can communicate their thoughts clearly, are more convincing in showing their disciplinary reasoning skills than students who cannot. It is thus an advantage for students to be able to diligently transform thoughts into text, which in itself should already be a solid reason to include disciplinary literacy in the classroom.

Moreover, the act of writing might enhance the learning process. Several studies have suggested that writing can become a powerful means of retrieving, rethinking, revising, and reformulating what one knows (e.g., Galbraith, 1999; Graham et al., 2020; Klein, 1999). In current practice, writing assignments are not always designed as writing-to-learn tasks, due to a lack of knowledge about the kinds of learning that writing might engender (Newell & Winograd, 1995). When a specific task is well considered, however, a writing assignment might serve this additional goal of learning, which is the most prominent objective of content classrooms.

1.2 Historical Writing

Writing in history is an eminently useful learning activity. Wiley and colleagues (2014) defined learning history as the attempt 'to try to understand the past'. Historical reasoning and problem solving often center on texts. 'Understanding the past' might thus be interpreted as the construction of mental models of past phenomena, based on various documents representing diverse perspectives. This is a learning goal that writing could support perfectly.

Historical writing entails that students must use historical evidence, drawn critically from primary source documents, to write well-structured and well-substantiated arguments (De La Paz & Felton, 2010). It should be distinguished from generic argumentation, as historical arguments require "conceptual understanding, procedural knowledge of historical analysis, an underlying grasp of the topic and discipline, and background content knowledge" (Monte-Sano, 2010, p. 560).

According to Geisler (1994), expertise in a certain discipline requires familiarity with its content and its rhetoric. For historical writing, novice history students may see historical knowledge as random bits to be reproduced, summarized and written down in a preconceived format to relate what they know about the topic. In contrast, expert historians would see historical knowledge as a construction of an evidenced interpretation of an issue, and use the rhetorical strategies of the disciplinary genre to transform contrasting bits of information into a coherent text (McCarthy Young & Leinhardt, 1998). This framework is consistent with Scardamalia and Bereiter's (1987) groundwork on novices and experts in writing (knowledge telling versus knowledge transformation). In upper secondary education, students will most likely not reach an expert level, especially when it comes to rhetoric, since this aspect is largely underexposed in history education when compared to content (Geisler, 1994).

1.3 Learner Characteristics

When designing an instructional unit aimed at historical writing, it should be taken into account that two learner characteristics might have a moderating effect. Firstly, students with high writing self-efficacy, generally show low writing apprehension and enjoy writing more, which is associated with higher text quality (Pajares & Valiante, 1997; Sanders-Reio et al., 2014). Secondly, students may hold implicit writing beliefs which relate to writing quality (White & Bruning, 2005). These beliefs comprise what students think proficient writing is, and what they think good writers do; these beliefs thus modify what their composing process will look like, and how the eventual text will

be established (Graham et al., 1993). White and Bruning (2005) identified transmissional and transactional beliefs. Writers with high transmissional beliefs generally view writing as a means for reporting 'someone else's facts'. These writers stay close to the information and arguments they find in source materials. In contrast, those writers with high transactional beliefs, are cognitively and emotionally engaged in their writing processes. They see writing as a means of developing their understanding of the issues at stake, and their own views on these issues. In our instructional unit, we aimed to appeal to students with all self-efficacy levels and all writing belief profiles.

1.4 Research Questions and Design Research Methodology

Following McKenney and Reeves' generic model for design in research (2019), we conducted a design study that sought to establish design principles for the integration of disciplinary writing into history, and the development of sound lesson materials for history teachers. For high quality interventions, Nieveen (1999) proposed four generic criteria: content validity, construct validity, practicality, and effectiveness. Our main research question was therefore:

What is a valid, practical and effective design for writing tasks, and writing instructions, for upper secondary school history to support students' historical writing?

McKenney and Reeves distinguished three phases in design research. During the analysis and exploration phase (1), a literature review and an explorative preliminary context study were conducted. During the design and construction phase (2), the design principles were formulated, and a conceptual model was created. Thirdly, the evaluation and reflection phase (3) involved a trial intervention study. Our research questions for the different phases were:

RQ phase 1: How is students' writing in history currently addressed by history teachers, which design requirements can be derived from these findings, and which design principles can be derived for approaching historical literacy? (i.e. focus on validity)

RQ phase 2: How can these design principles be translated into an instructional unit? (i.e. focus on validity and practicality)

RQ phase 3: How do students and teachers interact with the instructional unit, in what ways is this different from intended interactions,

with what results on historical writing, why does this seem to be the case, and what is the role of learner characteristics (writing beliefs and self-efficacy)? (i.e. focus on practicality and effectiveness)

In this work, we will report on the method and outcomes for each of these three phases sequentially. However, it must be noted that the design process was iterative, and thus the three phases interacted.

2. ANALYSIS AND EXPLORATION PHASE

To explore the context, we interviewed ten history teachers from different schools in the Netherlands (age ranged from 32 to 63 ($M = 42$)). These teachers were all qualified teachers (masters' degree) of upper secondary levels (years of experience ranged from 4 to 35 ($M = 16$)). Each interview took about one hour, and all was audio-taped and subsequently transformed into a written protocol. The aim of the interviews was to obtain a clear view of teachers' current writing and support practices.

We used a stimulated recall strategy: teachers were asked to bring a writing task, containing multiple sources, which they had used recently in an upper secondary class, and two example texts (one weak and one strong) from their students. This task, with accompanying example texts, was used as an angle for the interviews. We discussed five topics: task characteristics, assessment criteria, support practices, cognitive processes, and writing beliefs. For cognitive processes, we asked teachers to explain which processes students should perform, to complete the assignment successfully. Teachers wrote these processes on separate sticky notes, to subsequently paste and organize them on a piece of paper, and draw the connections between the processes. The interview guide can be found in Appendix E. The written protocols were analyzed per theme, with a specific lens for design requirements and principles, using a content analysis procedure (Krippendorff, 2013).

To discover how historical writing was addressed in previous studies, and with which results for historical writing, we conducted a specific literature search. This literature might guide us to formulate design requirements and design principles, which could aid our instructional design. These principles are generally considered to be the core of the instructional unit, and are therefore described to profoundly contribute to theoretical understanding (Rijlaarsdam et al., 2017). We searched for English-language journal articles describing writing interventions in history at secondary school levels (e.g., De

La Paz, 2005; De La Paz & Felton, 2010; De La Paz et al., 2017; Graham & Perin, 2007; Martinez et al., 2015; Monte-Sano, 2011; Reynolds & Perin, 2009; Van Drie 2005, 2014; 2015).

2.1 Interview Results

In this section, the interview results which were viewed most relevant for our design are concisely presented. First of all, with regard to support practices, our results confirmed findings in previous studies (e.g. De Oliveira, 2011): teachers rarely mentioned writing support practices, implying that writing support was not often provided, at least not consciously. Four out of ten teachers did not mention writing support practices at all. While the other six teachers did, this was mostly related to the intended text; for example, teachers explained how the eventual text should be structured, or which criteria were considered to be important. Three teachers did mention practices related to the process, for example, discussing how to assess source texts, or modeling how to deduce arguments from source texts. One of these three teachers showed awareness of the relevance of process instruction:

"If there's a source text, you read the question first. You then read the question again, and then you start thinking: what is being asked? And only then do you look in a source for the answer." (Teacher 10)

However, if support practices were mentioned, they focused on aspects of reading and analyzing source texts, and not on writing.

Regarding cognitive processes, we found that teachers rarely addressed any aspects of the writing process. Six teachers mentioned the writing process in terms of "writing" or "writing competence". One teacher, who had learned about writing discourse in a professional learning community, named the writing part of the process more extensively: for example, they mentioned "dividing into paragraphs" and "critical rereading and fine-tuning" as processes. Three teachers mentioned no steps focused on writing at all; they stuck strictly to historical skills, such as "analyzing sources" and "drawing conclusions," leaving writing implicit.

Nevertheless, all teachers mentioned writing skills as important for the subject of history. Several teachers noted that what students meant often differed from what they actually wrote. In such cases, students had generally not mastered the content convincingly. Teacher seven elaborated: "If a text is well written, then you immediately believe that the student understands it".

Moreover, five out of ten teachers mentioned the improvement of students' writing as a goal of the writing task they brought to the interview. Two of them considered "learning to express oneself" the main goal; for the other three teachers, writing development was a subgoal next to content knowledge goals. The remaining five teachers mentioned solely content goals as the learning goals for the writing task, and not writing goals.

2.2 Design Requirements

Several design requirements were derived from the exploration of the context. The interview results showed that teachers struggled to unequivocally explain which processes their writing tasks implied for their students. A first requirement for the design, therefore, was to consider the fact that history teachers lack knowledge about writing processes.

A second requirement, was to keep practicality in mind. Writing instruction need to be easily applicable for history teachers; interview results showed that teachers do not regularly provide students with writing instruction, or support during the writing process (also: De Oliveira, 2011). As teachers commonly find it difficult to teach reading and writing strategies, the instruction must be made as easily integrable as possible, for them to maximize practicality. Practicality can be divided into three criteria, all related to classroom ecology and a teacher's goal system (Westbroek et al., 2020). A first criterion is instrumentality: are procedures available to show how the innovation should be implemented? A second criterion is congruence: is the innovation sufficiently congruent with regular practices and important goals, that the teacher connects with? And as a third criterion, low cost: do the estimated benefits of implementation outweigh the effort it takes to implement the innovation? Keeping track of these practicality criteria is thus strongly recommended for the instructional design.

A third requirement centers around the finding of previous research that there is an ongoing subject culture, with 'content first' as an important element (e.g. O'Brien et al., 1995). Specifically, half of the interviewed teachers not mentioning "writing development" as a learning goal for a writing task, might be seen as a confirmation of this persistent culture. Although our aim was to eventually improve students' historical literacy, an important design requirement thus would be that literacy is developed in service of content.

2.3 Design Principles

A literature search to discover how previous studies addressed students' historical writing was conducted, including a search for intervention studies in secondary grades regarding historical writing combined with learning. De La Paz and colleagues (De La Paz & Felton, 2010; De La Paz et al., 2014; De La Paz et al., 2017) had similar objectives in their respective studies, and they reached encouraging results. Explicit instruction in disciplinary thinking which is embedded in reading-writing tasks, based on sources, supported students' historical argument writing. We have derived two design principles from the aforementioned studies by De La Paz, which are further presented and discussed in the subsequent section.

2.3.1 Design Principle #1: Writing Task Design

Similar to De La Paz and colleagues, we aimed students to learn through writing. In their studies, students wrote 'historical inquiries', which they explain as 'working with conflicting primary sources to investigate a central question' (De La Paz et al., 2017, p. 36). Several authors (Monte-Sano & De La Paz, 2012; Newell & Winograd, 1995; Van Drie, et al., 2006; Voss & Wiley 1997) have previously explored the task effects on learning; these studies have underscored that the arguing genre is best suited for writing to learn, since this genre will promote deeper processing of content materials.

Van Drie and colleagues (2006) concluded that evaluative prompts (e.g. 'Were the changes in the behaviour of Dutch youths in the 1960s revolutionary?') were most powerful for eliciting historical reasoning, compared to an explanatory prompt (e.g. 'How can the changes in the behaviour of Dutch youths in the 1960s be explained?'). Our think-aloud study (Chapter 2) showed that with such evaluative tasks, the writing process itself is likely to enhance students' thought process, which subsequently led to inclusion of this specific type of argument tasks into our design.

To enhance learning, it is advised to use source-based writing tasks. The sourcing skill is considered key in history education (Wineburg, 1991), which makes source-based writing a suitable learning activity (Brante & Strømsø, 2018). This was confirmed by a study of Monte-Sano and De La Paz (2012), which revealed that writing tasks which were stimulating engagement in sourcing, corroboration, and causal analysis, improved students' capability of recognizing and reconciling historical perspectives significantly. Furthermore, as Voss and Wiley (1997) concluded, writing arguments from sources might facilitate understanding of content.

Writing tasks should contain multiple documents representing multiple perspectives on the issue at stake, since documents are written from an author's perspective, and no single perspective can be an exhaustive presentation of a historical phenomenon (Britt & Rouet, 2012). Additionally, inclusion of primary documents is recommended, since these stimulate students' attention to source characteristics and trustworthiness, and make students more likely to include references to source documents in their own writing (Mierwald et al., 2022; Rouet et al., 1996). After all, constructing arguments from several primary sources, is the essence of history; students learn that history is not just about learning names and dates. It is "an on-going debate about what those facts may mean" (Voss & Wiley, 1997, p. 264).

Furthermore, the writing tasks should not result in extensive texts, in order to avoid the risk of discouraging teachers by an overwhelming paper load (Newell & Winograd, 1995). Hence, it is best to use tasks that can be performed within one lesson, resulting in texts of approximately 200-300 words, which is much shorter than the tasks and texts in other studies. Lastly, frequent practicing is preferred (Graham & Harris, 2017), which is also easier achieved with shorter tasks. To summarize, our first design principle is therefore:

Design principle #1: If we want students to develop a profound understanding of history through writing, then they should write short evaluative texts, based on multiple primary sources which represent multiple perspectives.

2.3.2 Design Principle #: Writing Process Support

A second key element in our design, considers teachers' support of students writing. From our exploration of the context, we concluded writing support was absent, or merely focused on the product. Therefore, discipline specific strategy instructions should be included, to support students during writing. This strategy instruction should be discipline-specific, since this is preferable exceeding simply adding a general reading-writing strategy instruction to the disciplinary classroom (Gillis, 2014).

The writing part of the reading-writing process should be emphasized, since this part of the process was most disregarded by the interviewed teachers. Monte-Sano and Allen (2018) drew a similar conclusion, as history teachers from their respective studies tended to focus more on the historical work involved in writing, than on the composition of text. However, a design

requirement for the strategy instruction, is that it could be implemented by teachers with no, or only basic knowledge, of writing processes.

Furthermore, it is recommended that the instruction should be flexible, in order to match students' writing preferences (Van Ockenburg et al., 2021b), and adaptable to the task, and task difficulty (Chapter 2). Students who have more background knowledge, tend to move faster through the process, and start writing sooner in the process compared to students who have less background knowledge (Chapter 2). The latter generally stay longer in the phase of reading and planning. Students' writing routines are likely to differ from task to task, and from topic to topic (van Steendam et al., 2022). Furthermore, research on writing-to-learn about literary texts, has indicated that adapting writing tasks to students' writing strategies, increases their learning in the field of literature (Kieft et al., 2008). Therefore, we will offer students a dual-route strategy. These two aspects – focus on writing and flexibility – of the instruction are distinctive to our design. To summarize, our second design principle is:

Design principle #2: If we want students to develop a profound understanding of history through writing, it is best to provide them with discipline-specific, dual-route, reading-writing strategy instruction which is easily applicable for teachers.

3. DESIGN AND CONSTRUCTION PHASE

In the second design phase, we developed the design principles into a prototype, considering the aforementioned design requirements. The design-as-constructed is presented, and supported by literature, in the following section.

3.1 Design-as-Constructed

The first design principle, has implied the development of a task intervention. As learning activities, students perform short evaluative writing tasks. The second design principle had implied an instructional intervention, since we want to support students' performance when executing this type of tasks. We have developed a discipline-specific strategy as the object of the instruction. Previous intervention studies on instructional design provided our framework for development of the strategy, and for the instructional design. We will present these two aspects of the intervention in the next sections.

3.1.1 *Development of the Strategy*

To decide which strategy was most suited, we conducted a literature analysis on earlier intervention studies which evaluated reading-writing strategies in history. Several studies have evaluated strategies for different parts of the reading-writing process (e.g. De La Paz, 2007; Martinez et al., 2015; Montesano, 2011; Reynolds & Perin, 2009). De La Paz (De La Paz & Felton, 2010; De La Paz et al., 2017) evaluated an integrated approach: reading, writing, and history, were taught together, with separate strategies for each phase. In our design, we combined strategies for reading, writing, and historical reasoning, into one overall Read-Think-Write strategy, based on previous literature on reading processes, writing processes, and disciplinary reading and writing. The strategy entails seven steps, which we will substantiate in the following section. The strategy as presented to students, can be found in the supplementary materials.

Monitoring. According to Britt and Rouet (2012), evaluative questions require students to coordinate a series of somewhat iterative steps. Students constantly need to change roles: from reader-thinker, to thinker-writer, to writer-reader. These roles are closely interrelated, necessitating flexible processing of each step (Britt & Rouet, 2012; Rouet & Britt, 2011). Students need to become aware of those role-switches, which can be established by means of monitoring prompts.

Reading. At the start of the process, students explore the task and sources. According to Britt and Rouet (2012), students should establish what they already know, and what they need to know, in order to write a response to the question. Furthermore, students should have a functional understanding of the genre requested in the prompt. Step 1, analyzing the task, is thus about constructing a task model (Britt & Rouet, 2012). The task model includes a) the task goal; b) actions to achieve the goal; and c) a set of criteria for reaching the goal. In our study, the task goal is to write an evaluative text that takes a stand on a particular controversy, which is supported by discipline-appropriate evidence, but also deals with perspectives that are contrary to the student's stance. At a certain moment during the phase of building a task model, students begin to turn this task goal into a set of actions, in order to establish an educated stance on the issue. Such actions consist of activities such as reading sufficient source information, identifying potential supporting reasons and evidence, selecting strong reasons with respect to available evidence and

audience, identifying other-side positions which must be addressed, and identifying appropriate responses or rebuttals for those counterarguments, if possible. Finally, students should identify criteria for accomplishing task goals. To properly turn the task specifications into goals and action plans, students must understand the nouns ('colonial art', 'western museums') and the action words ('should', 'write an argument').

The next step, Step 2, is to analyze source materials. For each source, students should generate a situation model (Britt & Rouet, 2012; Rouet & Britt, 2011). This model begins with the assumption that texts are experienced as social entities, not simply as a series of linguistic propositions (Wineburg, 1994). Therefore, it is important to detect several features beyond propositional content, such as the author, the genre, the publication date, the intended audience, and the purpose. Understanding of each source text is necessary, to continue the process.

Thinking. Students gradually shift their role from reader to thinker, and to some extent, to writer. Step 3 is to connect the source information: in this phase an intertext model is created: readers generate links between sources, which can include rhetorical relations such as corroborating, supporting, or contrasting (Britt & Rouet, 2012). Eventually an integrated mental model of the situation or phenomenon is created: an internal representation of the situation or phenomenon described across texts (Britt & Rouet, 2012). The structure of the information in this mental model will depend on the content and the nature of the task. The critical point here, is that in a multiple document situation, the reader of the sources is the author of the integrated mental model. This generally requires the content to be transformed and to be re-organized. For Step 4, students are advised to reread the question to keep track of the focus of the question. This step might be seen as a monitoring prompt.

At Step 5, students begin to answer the question. Previous research has shown that we can distinguish different sorts of writing processes (Torrance et al., 1994). Kieft et al. (2008) recommend that writing instruction is adjusted to individual writing strategy preferences. Therefore, in our strategy two routes are amplified in this phase, both well suited for evaluative history tasks: (1) the route of writing freely, where students write down all they can think of in a few writing spurts; and (2) the route of pre-planning, where students build their text based on a text scheme and use many short writing spurts (Chapter 2).

The student should have a clear picture of the main structure of the text by now. An evaluative question generally requires a dual-perspective argumentative text, complemented by a position based on a consideration of all arguments.

Writing. The steps in the final phase concern communication. Students will need to continuously switch roles again, from writer-thinker to reader, and back. We distinguish between revising (Step 6) and editing (Step 7) during this phase. Flower and Hayes (1981) argued that the goal of revision is substantive change: 'revision can lead to re-seeing, restructuring, even reconceptualizing the entire discourse' (p. 16). Revision could take place at word level, sentence level, or paragraph level. Especially, free writers will need to invest in extensive revision in order to transform their draft version into a communicative text. The pre-planners will mainly expand their text, from the text plan they developed. The final step for all types of writers, is to edit for language use, minor errors on word or sentence level, and typos.

3.1.2 Instructional Model and Key Learning Activities

Previous studies on intervention research are quite unified in their recommendations for instructional design, as many studies have used the 'classic' model of Self-Regulated Strategy Development (SRSD) (Harris & Graham, 1996) as a base. This model consists of six instructional stages with gradual release of support: (1) develop background knowledge; (2) discuss it; (3) model it; (4) memorize it; (5) support it, and (6) independent performance. Several intervention studies aimed at writing-to-learn history have used SRSD, albeit in slightly adjusted versions (e.g. De La Paz & Felton, 2010; De La Paz et al., 2017; Martinez et al., 2015; Reynolds & Perin, 2009). In our study, we followed the model of De La Paz and Felton (2010), who themselves based their design on the cognitive apprenticeship approach to instruction, as developed by Brown et al. (1989). Five stages provided the framework for instruction: develop background knowledge, describe it, model it, support it, and independent performance, which was filled in with learning activities (Table 3.1).

Table 3.1. Description of the Design-as-Constructed

Lesson*	Stage	Learning activities	Description
T1		Pretest	Prompt: To what extent have the United Nations been successful in the past 50 years?
0	<i>Develop background knowledge</i>	Task experience	Students perform an evaluative writing task to 'experience' the task. Prompt: To what extent do you think the Dutch government should return colonial art stolen from local communities in Indonesia during Dutch colonial rule?
1		Reflection	The students write down what they thought was easy while performing the task on colonial art, and what was difficult.
	<i>Describe it</i>	Building new knowledge	The teacher presents and explains the RTW strategy.
	<i>Model it (process)</i>	Comparing strategy to own experience	The students compare their own experiences with evaluative tasks to the presented strategy.
		Observing strategy demonstrated by a modeling peer (video)	Students watch a 12-minute video in a plenary session on the main screen. This video presents the strategy and contains fragments showing a modeling peer, who demonstrates how each step of the strategy could be performed.
		Assessing a peer's performance	As a processing activity, the students individually compare the performance of the modeling peer with the strategy as presented by scoring the peer on a scale from 0 to 100.
2	<i>Model it (product)</i>	Assessing peers' texts	Students individually assess three model texts (on the colonial art-task, lesson 0).
		Generating criteria	In a class discussion, students generate a criteria list.
		Applying new learning	Students apply the criteria by revising the text they wrote about colonial art (lesson 0), with the criteria list in mind.
3/T2*	<i>Support it</i>	Scaffolded practice	Prompt: To what extent did the images the United States and Soviet Union had of each other play a role in the Vietnam War? Students perform a second evaluative writing task, scaffolded by the support of the teacher and the written guide (strategy cheat sheet). Students are encouraged to collaborate in prewriting.
4		Teacher feedback	The teacher provides the class with feedback on the task, as he is used to do with similar history assignments.
5/T3*	<i>Independent performance</i>	Individual work with the help of a written guide	Prompt: To what extent were American and Vietnamese citizens involved in the Vietnam War? Students perform a third evaluative task based on sources with the help of the written guide (strategy cheat sheet).
6		Teacher feedback	The teacher provides the class with feedback on the task, as he is used to do with similar history assignments.

Key Lessons: Strategy Instruction

*The design-as-constructed was implemented as presented in Group 1 only. In Group 2, the key intervention lessons (Lesson 0-1-2) were provided after T2 performance (see also Figure 3.2).

Note: Lessons of the instructional unit were interspersed with content lessons (except for Lesson 0-1-2, which were taught consecutively).

We developed a series of learning activities, to be spread over seven 50-minute lessons. A paper workbook was developed for all students, which contained each assignment.

The instructional unit starts with an experience of the task in Lesson 0. In this Lesson, students perform an evaluative task on colonial art. The central issue of this task was: 'To what extent do you think Western museums should return colonial art to the country of origin?' Four sources were provided, all textual (mean length: 189 words). Multiple perspectives were represented, from: (a) a Dutch assistant resident in Indonesia (in 1856); (b) a Dutch artist who accompanied the army on a military expedition to Java, to collect Javanese art (around 1906); (c) the current king of Klungkung, Bali (in 2020); and (d) the director of the National Museum of Indonesia (in 2020).

In Lesson 1, students were prompted to reflect on their writing process. Subsequently, they were presented the Read-Think-Write strategy by direct instruction, and they watched a video showing a modeling peer, who demonstrated each step (Figure 3.1). Fragments of modeling were alternated with instructional parts, explaining the strategy. The modeling peer was a volunteering 12th grade student, who was pre-informed about the strategy and instructed to demonstrate all steps while performing the task about colonial art from Lesson 0. It took two sessions of about 40 minutes of filming, to eventually create a video of 12 min 52 s, in which the modeling peer was visible and audible during 7 min 38 s. While watching the video, students performed a compare-contrast assignment to help them reflect on their writing process.

In Lesson 2, students discussed exemplars, which were printed in their workbooks. These texts were collected in an earlier stage and written by 11th grade students in response to the task of Lesson 0. Analyzing and discussing exemplars is a means of imparting teachers' tacit knowledge of criteria (Polanyi, 1973). Discussing exemplars has been shown to reduce differences between student and teacher ratings in previous research (Orsmond et al., 2002). The learning activity thus contributes to task representation.

Lessons 3 and 5 contained writing tasks developed by students' teachers, tailored to the regular curriculum. These tasks were also included in the paper workbook, but students were instructed to write their texts on a computer. Similar to Martínez et al. (2015), we added a written guide ('Cheat sheet', see supplementary materials) as an additional support in the support it and independent performance-stages (Lessons 3-6). In Lesson 4 and 6, the teachers were considered to provide feedback on students' texts.

Figure 3.1. Video Fragments: An Instructional Slide (on the Left) and a Still From the Modeling Peer Illustrating the Process (on the Right). The Original Video Was in Dutch.



In the subsequent phase of the study, we implemented the design-as-constructed into practice, which is described in the next sections.

4. EVALUATION AND REFLECTION PHASE

In the third phase, we implemented the instructional unit into the practice of two history teachers. Method and results of this trial intervention study are presented in the next sections.

4.1 Research Design

For the implementation, we used a switching panels design, with three measurement occasions (T1-T2-T3). An overview of the research design is presented in Figure 3.2. Teacher A implemented Lesson 0-1-2 from the instructional unit between T1 and T2, teacher B between T2 and T3. In between lessons of the instructional unit, lessons focused on knowledge building. Content lessons were also provided, in order to prepare students for the historical issues discussed in the writing tasks.

The three measurements were spread over a period of 8 weeks. For ethical reasons – we did not wish to waste lesson time purely for testing –, we integrated our measurements into the regular lessons. In Group 1, therefore, T2 was Lesson 3, and T3 was Lesson 5. In Group 2, T3 was Lesson 3. Lessons 5 and 6 were absent in this Group.

The switching panels design provided us with the opportunity to implement the intervention lessons twice in a row (first in Group 1, then in Group 2), with the groups functioning as each other's control group. Moreover, students were not withheld writing instruction with this design, as is ethically preferable (Cohen et al., 2011).

Figure 3.2. Overview of the Research Design

Panel 1					Panel 2					
Group 1 Teacher A <i>n</i> = 31	T1	History lessons	Intervention	History lessons	T2	History lessons	Lesson 4 / Feedback lesson	History lessons	T3	Lesson 6 Feed-back lesson
			Lesson 0-1-2 Strategy instruction		Lesson 3 Application of strategy				Lesson 5 Application of strategy	
Group 2 Teacher B <i>n</i> = 30	T1	History lessons	History lessons	History lessons	T2	History lessons	Intervention	History lessons	T3	Lesson 4 Feed-back lesson
			History lessons		Lesson 0-1-2 Strategy instruction		Lesson 3 Application of strategy			

Note: "Lesson" refers to the lessons described in Table 3.1.

4.2 Participants

Two qualified teachers of the same school implemented the intervention lessons in their 11th grade (16-17 yrs old, pre-university track) history classes. Teacher A (a 58-year-old male) had 33 years of experience teaching history, while Teacher B (a 42-year-old female) had 17 years, respectively. Neither teacher had any experience in teaching writing. The class of teacher A consisted of 31 students (20 female); while teacher B taught 30 students (11 female). The students were all actively informed about the research procedure, and gave passive consent for participation.

4.3 Measurements

4.3.1 *Practicality: Reflective Interviews*

We aimed to gain insight into the functionality of the intervention, by reflective interviews with both teachers and students. The interview guides are added in the supplementary materials. We conducted reflective interviews of approximately 45 minutes with each teacher. We also conducted two 45-minute group interview sessions with students, one session per group, and each time with two students (Group 1: Student A1 and A2; Group 2: Student B1 and B2). The students were all female, and participated voluntarily. All interviews were conducted online (via Microsoft Teams), and audio-recorded, for which all interviewees actively consented. The audio-recordings were subsequently translated into written protocols.

4.3.2 *Effectiveness: Text Quality*

Writing Tasks. To indicate effectiveness, we used three writing tasks which were co-designed with the participating teachers. The tasks fitted into the regular curriculum, and replaced regular workbook assignments. Students developed basic content knowledge on the topic of the tasks, before executing each task. The topic of the term, was the Vietnam war. Task 1 focused on the question "To what extent have the United Nations been successful in the past 50 years?". Seven sources were presented: two cartoons, and five textual sources (mean length: 183 words). Task 2 centered around the question "To what extent did the images the United States and Soviet Union had of each other play a role in the Vietnam war?". Six sources were presented: one cartoon, and five textual sources (mean length: 129 words). Task 3 queried "To what extent were American and Vietnamese citizens involved in the Vietnam war?". Five sources were presented, all textual (mean length: 180 words). At all three measurements, students were also allowed to use textbook materials.

All tasks contained sources which represented multiple perspectives, and were predominantly primary.

We used text quality as an indicator of effectiveness on three dimensions: (1) quality of content, (2) quality of structure, and (3) holistic quality. We used rubrics for the assessment of content quality and structure. The rubric for content quality was based on the studies of De La Paz and colleagues (2017). They proposed four substantive criteria for students' historical texts – contextualization, rebuttal, substantiation, and perspective recognition – which we adopted, and complemented with additions from several other intervention studies (Van Drie et al., 2015; Martinez et al., 2015; Britt & Aglinskis, 2002; Voss & Wiley, 1997; McCarthy Young & Leinhardt, 1998). This resulted in an operationalization of the content criterion into the following four aspects: (a) understanding of the issue, (b) multi-perspectivity, (c) elaboration of argumentation, and (d) use of sources.

For quality of structure, we focused on the organizational pattern of the text, a decision that was based on previous studies of McCarthy Young and Leinhardt (1998) and Reynolds and Perin (2009). De La Paz (2010, 2017) did not include structure as a separate criterion; however, their holistic scoring rubrics highlight rhetorical aspects, also with a focus on overall text structure and coherence. The rubric contained two aspects: (a) global and (b) local text structure. The complete rubric for content and structure, can be found in Appendix B.

For assessing students' texts holistically, we constructed a scale with five example texts, exemplifying different levels of quality. For the selection of these five texts, we went through four steps. As a start, one author (LH) selected 30 texts written by students out of a pre-collected pool of 55 texts about the colonial art task; we ascertained that the sample covered three levels of quality: weak (20%), average (60%), and good texts (20%). Subsequently, a random jury panel of 11 teachers was asked to rank these 30 texts using *Comproved*, an online tool for comparative judgment (<https://www.comproved.com>). Each of these teachers was asked to make 30 comparisons between randomly composed pairs of texts, and to decide which text was better. For each fifth comparison, we asked for an explanation of the teacher's preference. These elaborations gave us the opportunity to refine the scale with annotations which could explain quality differences.

The third step, was to select five texts, which were evenly distributed in the rank order that was constructed by the jury. We selected the texts with z-scores closest to -2, -1, 0, +1, and +2. Additionally, we checked if standard

errors of these texts were not exceptionally large. This resulted in the selection of the texts from ranks 3 ($z = +1.88$, $se = 0.59$), 6 ($z = +1.07$, $se = 0.52$), 15 ($z = -0.01$, $se = 0.49$), 24 ($z = +0.98$, $se = 0.55$), and 29 ($z = +1.88$, $se = 0.61$), to function as example texts in the scale. The final step was to annotate the scale on the criteria content and structure.

4.3.3 Learner Characteristics: Writing Beliefs and Self-Efficacy Questionnaires

The learner variables were measured by two questionnaires (Appendix C). The validated writing beliefs questionnaire (Vandermeulen et al., 2020) consisted of 26 statements to rate on a five-point Likert scale. We distinguished two scales: a transmission scale ($\alpha = .68$) (e.g., "I write to clarify what others think") and a transaction scale ($\alpha = .74$) (e.g., "Writing helps me see the complexity of ideas") (White & Bruning, 2005). The two beliefs did not correlate ($r = -.08$, $p = 0.61$).

The self-efficacy questionnaire consisted of 30 statements to rate on a scale from 0 ("I cannot do this at all") to 100 ("I can do this perfectly"). We constructed this self-efficacy questionnaire based on a previous study (Chapter 2), covering ten aspects involved in historical writing tasks: reading and analyzing sources; connecting information; selecting information; synthesizing; use of sources; content; coherence; structure; language, and monitoring. These aspects were not intended to be separate constructs; we intended to cover the whole reading-writing process as it is involved in an evaluative task. The Cronbach's alpha of the self-efficacy questionnaire was .95. The mean score of the 30 items functioned as self-efficacy score.

4.3.4 Fidelity Measures

To check whether the implementation of the intervention was indeed executed as intended, we used O'Donnell's (2008) five criteria for measuring fidelity of implementation. The first criterion is (a) adherence; to check whether the components of the intervention were delivered as designed, the teachers completed teacher logs via email. After each interventional lesson, the teachers updated the team on what was done in the lesson, what plans they had for the next lesson, and what adjustments they had made in the lessons as designed, and in their plans. We also conducted reflective interviews, and one of teacher A's lessons was observed by LH. In the reflective interviews with teachers and students, we also checked for (b) duration – the number, length, and frequency of implemented lessons, (c) the quality of delivery, (d)

participant responsiveness, and (e) program differentiation – whether critical features that distinguish the program from the comparison condition were present during implementation.

4.4 Data Collection

4.4.1 Procedures and Circumstances during the Intervention Period

Unfortunately, the research was conducted during a period of several (partial) CoVid-lockdowns, and therefore there were restrictions to adapt to. The project started while the country was in a lockdown situation, during which only fully online secondary education was provided; i.e. the teacher taught from home, via Microsoft Teams. After a few weeks, just before Task 2, the situation changed into a hybrid situation, i.e., the teacher taught from school, with half of the students present in the classroom, and the other half of the students attending online. The two groups (online-offline) alternated every full day, so students were physically at school five days out of every two weeks. Groups were split randomly. Group 1 received the strategy instruction in a full online setting; Group 2 in a hybrid situation.

During the regular knowledge building lessons, the teachers explained history content in a traditional way: the teacher explained the course, as well as causes of the Vietnam war. The teachers interacted with the students now and then, in order to keep them alert, but on the whole, the students mainly listened and took notes. This was alternated with short tasks, such as drawing timetables and discussing maps of the Vietnam war situation.

4.4.2 Procedures and Circumstances During Measurements

The teacher supervised the writing tasks and questionnaires, which were performed on a computer, and were handed in digitally, using an electronic learning system. All students performed the questionnaires and Task 1 online, while at home. Tasks 2 and 3 were performed in the hybrid situation, with half of the students at home, and the other half in a computer classroom. Supervision varied in the online situations; students who were at home were not always willing to put their cameras/microphones on. Most texts (80%) were written on the computer, some texts were written by hand (20%), photographed, and handed in digitally. The number of handwritten texts was larger in Group 1 (31% of the total number of texts was written by hand) than in Group 2 (6% handwritten texts). For the analysis, all handwritten texts were transformed into typed texts.

4.4.3 Rating Procedures

The data set consisted of 121 texts (T1: 49, T2: 36, T3: 36) with three different topics, which were randomly divided per topic, over seven text sets (each containing the three topics; approximately 17 texts per topic). A panel of seven raters evaluated the texts: all raters were qualified current or former teachers of history. Each rater rated three different text sets (total approximately 50 texts), partially overlapping with four other raters, resulting in all texts receiving three rates by three different raters for all three variables. The intraclass correlation coefficients of these three measures were .79 for the holistic score, .73 for content, and .77 for structure. Further analyses were conducted using the means of the three ratings.

4.4.4 Implementation Fidelity

From the fidelity measures, several deviations from the design-as-constructed were observed. Firstly, from the interviews, we noted that for teacher B, the model texts remained underexposed in the implementation, potentially since they did not feel confident teaching this particular assignment. Students in Group 2 did examine the model texts, but they were then not further discussed with the teacher (as occurred in Group 1). The quality of delivery might thus have been lower in Group 2.

Secondly, feedback lessons in both groups were short (feedback took about 10 minutes of the lesson), and these were not focused on students' writing processes. The teacher manual did not provide strict guidance here; teachers were free to discuss the tasks at their discretion. They were encouraged to refer to the strategy whenever possible, but in practice, teachers only discussed content. For example, they highlighted which elements in the text were necessary to discuss the task topic properly.

We also signaled some deviations from the design-as-constructed that can be directly linked to the teaching being in an online setting. For instance, online education impeded interaction with students. For example, the practice task from Lesson 3 (see Table 1) was intended as a scaffolded task, but because of the online education, students performed the writing task mostly individually, without much scaffolding from peers or their teacher. Although it was possible to consult the teacher, practice showed us that students hardly asked questions in online situations. Furthermore, the individual periods were shortened from 50 to 40 minutes quite ad hoc. This caused teacher A to spread Lesson 1 from the intervention over two periods, to meet the teaching goals.

Moreover, participant responsiveness was only moderate. For example, in Group 1, 55% completed all three tasks, while in Group 2 this was 30%. The merit for students to hand in their work declined over the tasks: in Group 1 the response rate for T1-T2-T3 was respectively 84%, 65%, 70%, and in Group 2 this was 77%, 53%, 47%. A possible explanation for this decline, could be students' demotivation caused by the ongoing online class situation. In a short questionnaire, students were afterwards asked why they had not handed in their work in time. They self-reported several reasons: they said they had 'had no time for it', they had 'forgotten to do it', or they had 'not considered it a priority'. These remarks indicated that many students did not finish the tasks during the lesson time which was appointed to the performance of the task by the teachers. Both teachers noted in their evaluations, that the students' overall motivation declined during the lockdown period. and that it was a challenge to get students motivated in general.

4.5 Data Analysis

4.5.1 Analyzing Text Quality

Indicators for text quality were: (1) holistic quality; (2) quality of content; (3) quality of structure, and (4) text length. Specifically, text length was considered to be an indicator of text quality, as good writers generally write longer texts (Ferrari et al., 1998).

The scores on the criteria content, and on the structure, correlated substantially (T1: $r = .89$; T2: $r = .80$; T3: $r = .83$). Both structure ($r = .94/.89/.92$, $p < .01$) and content ($r = .93/.92/.89$, $p < .01$) also correlated strongly with the holistic score. Text length correlated with holistic quality, at .77, .82 and .87, respectively ($p < .01$). Despite these correlations, we nonetheless report further analyses for all indicators, as due to the small scale of this trial study, all possible influential factors should be examined.

For the analyses of students' scores, we conducted a multilevel analysis. Effects of the interventions in both panels were tested by comparing four nested models: Model 0 with Subject as random factor, Model 1 with the added factor Time, Model 2 with the added factor Condition, and Model 3 with the added interaction Time*Condition. We set the alpha level on $p < .10$, to avoid a Type II error in the case of a small sample. The comparisons of the fit of the nested models are presented in Appendix D.

4.5.2 Analyzing Effect of Moderator Variables

To explore moderating effects of learner variables (beliefs and self-efficacy), we first centered the scores for three potential moderators, and then extended the models presented above, with Model 4 (general effect of the learner variable LV); Model 5 (LV*Time); Model 6 (LV*Condition), and Model 7 (LV*Time*Condition).

Four students showed an outlier score (outside 95% the confidence interval of the mean) on one of the belief variables (high on Transaction 2x, low on Transaction 1x; low on self-efficacy 1x). These cases were not included in analyses in which belief-variables were explored for moderating a condition effect.

4.5.3 Preliminary Analyses

Prior to the intervention, the two groups did not differ with regard to beliefs and self-efficacy (Pillai's Trace .111, $F(3,40) = .434, p = .730$). Students' mean scores are presented in Table 3.2. The mean score for self-efficacy (70 out of 100) showed that on average, participants found themselves sufficiently capable for the type of tasks we asked them to perform.

To check whether data loss during the intervention was non-systematic in terms of the three belief scores, we ran multivariate analyses for T2 and T3. At T2 no effects of groups were observed (Pillai's Trace .069, $F(3,29) = .714, p = .552$). At T3, however, a multivariate effect was observed (Pillai's Trace .271, $F(3,27) = 3.353, p = .034$). Subsequent univariate analyses showed a significant difference between groups for transactional beliefs at Task 3; Group 1 scored significantly higher than Group 2 (3.4 vs 2.8, $F(1,29) = 9.499, p = .004$).

Table 3.2. Mean Beliefs and Self-Efficacy Scores

	Group 1 N = 23		Group 2 N = 21	
	M	SD	M	SD
Transmissional beliefs	3.19	.52	3.15	.43
Transactional beliefs	3.29	.46	3.11	.65
Self-efficacy	70.15	13.02	69.00	10.40

5. RESULTS

In this section, results of the trial intervention are presented on three criteria of high-quality interventions: validity, practicality and effectiveness.

5.1 Validity and Practicality**5.1.1 Teachers' Perspective**

Overall, teachers were positive about the tasks; they would surely reuse them in their teaching, also in other grades, because they felt that such tasks opened doors for a discussion of relevant historical issues they wanted their students to become aware of:

Teacher A: "For me, this is the essence of history, that you can have very different views on facts and events, based on factual material, and interpretations."

The teachers were less convinced they would reuse the strategy instruction, although they were both highly positive about the usefulness of the strategy itself, and they would re-use that in their lessons.

The video evoked divergent reactions. The modeling peer, and the explanation of the strategy, were useful elements, but they suggested that the duration of the explanatory part of the video could be shorter, in order to improve ease-of-use for students. Moreover, as the video was built upon the task regarding colonial art, they felt this to hinder transferability to other grades or topics. Teacher A was most positive about the video:

Teacher A: "I noticed the students were very interested to see how someone who is struggling, I think that's worth a million, to demonstrate those steps so visually."

Noticeably, both teachers seemed very selective in their enthusiasm about the possible reuse of intervention materials. Especially the aspects about the writing processes (Steps 5, 6, and 7 of the strategy) seemed to lack relevance in their opinion. They wondered whether the steps on writing might be shortened, or even be omitted completely.

Teacher A: "As a historian, I think: I have finished step 1-5 of that strategy, and that's enough, but there are another two steps on that sheet for some reason. Irritating."

Teacher B: "About the writing part, I thought: wow, that's pretty extensive. Could that be shortened? I believe so. I'm not sure, though."

5.1.2 Students' Perspective

Students perceived the writing tasks as useful and valuable, and they would like to do such tasks more often:

Student A2: "I liked the tasks. With workbook questions, it's often just a matter of reading comprehension. Questions are asked, and then it's just literally in the text. But now, you are really learning to think."

Although the tasks had increased the students' self-perceived knowledge, students did experience them as "a lot of work". With shorter tasks, with for example fewer source texts, students indicated they would feel less obstructed, or discouraged, by the estimated workload.

In the students' opinion, the strategy was very extensive; they thought the assignment would take too long, if they would perform all the steps. Some steps were "obvious": students felt they already knew them, and already did what was taught. Students liked to compare their own performance to the strategy, but they were not too excited about the video. The strategy was also explained on paper, so the explanatory part of the video felt redundant, and the modeling peer was too studious. To conclude: they thought of the video as 'dull' and 'slow'.

The model texts were appreciated; these provided the students with a good representation of what an evaluative text might look like, and the students said the model texts were helpful to become aware of the necessary components:

Student A2: "Through model texts, you learn which things are important in a text, and which things you should not do, which you otherwise might accidentally do yourself."

5.2 Effectiveness

For holistic quality, structure, and text length, we did not find interaction effects of Time and Condition, as we had expected. Model 3, which indicates such an interaction effect, fitted only the scores for content. Content score results are presented in Figure 3.3. However, upon exploring differences between groups at each measurement occasion, we found no significant results

(T1: mean difference = .153, $se = .244$, $p = .532$; T2 mean difference = .342, $se = .300$, $p = .262$; T3 mean difference = -.442, $se = .297$, $p = .146$).

Figure 3.3. Content Score Results (Scale 1-5) for Each Task (T1-T2-T3)

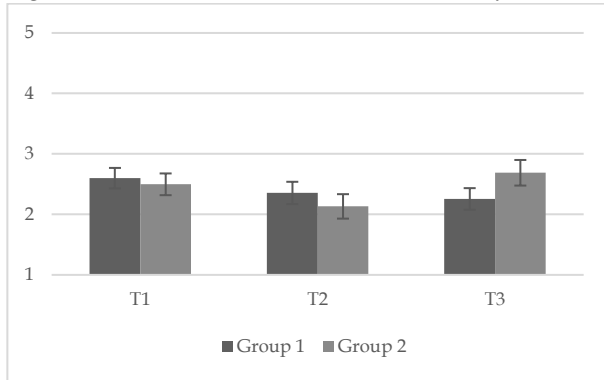
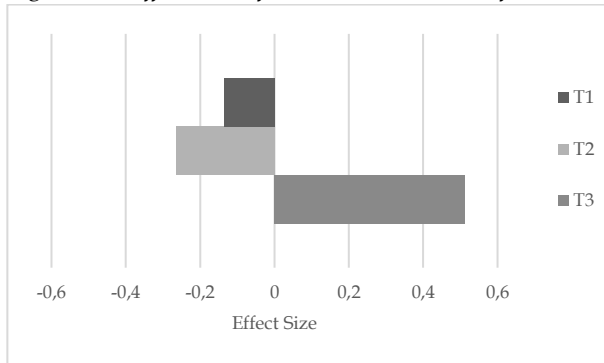


Figure 3.4. Effect Sizes for Content Scores (Reference = Group 1) for Each Task



To point out the scale of the differences between the groups at each measurement occasion, Figure 3.4 shows the effect sizes. At T3, the level of difference between the groups was .51, which might be considered a moderate effect size.

5.3 Moderating Effects of Learner Variables

We tested if one of the learner variables moderated the intervention effect. No moderation was revealed, for none of the three learner variables, and for none of the dependent variables.

What we did observe, with one exception*, was that adding learner variables in Model 4 resulted in a better fit (Appendix D) for self-efficacy, and for transactional beliefs, indicating direct effects of these factors. In all cases, the effect of self-efficacy was positive, varying from $\beta = .03$ ($p = .02$) for content, $.03$ ($p = .03$) for structure, and $.48$ ($p = .03$) for holistic quality. The effect of transactional beliefs was negative (content: $-.82$, $p = .004$; structure: -0.75 , $p = .02$; text length: -1.09 , $p = .002$; holistic quality: $-.15$, $p = .01$).

5.4 Effect of Online/Offline

In the second panel, a new variable was introduced: the hybrid situation, during which students were assigned into off- or online participation, by the school. For none of the four dependent variables, an effect of this instruction condition was observed.

6. CONCLUSIONS AND DISCUSSION

While every discipline has its own crucial specificities when it comes to reading and writing in upper secondary grades, the overarching question remains how best to support students in acquiring these skills. In the current study, we have described our instructional design process, to develop materials that would help teachers improve 11th grade students' historical writing. In this paper, we have reported on the design process, which contained three phases: (1) an analysis and exploration phase, (2) a design and construction phase, and (3) an evaluation and reflection phase.

In the first phase, we have explored the educational context through an interview study, resulting in three design requirements: it is advised to take into account (a) teachers' lack of knowledge about writing processes, (b) practicality demands, and (c) the subject culture of 'content first'. Furthermore, a literature search led to two design principles for our instructional design:

If we want students to develop a profound understanding of history through writing, then

Design principle #1: students should write short evaluative texts, based on multiple primary sources that represent multiple perspectives.

* The exception was self-efficacy in the analysis for text length.

Design principle #2: students should be provided with discipline-specific, dual-route, reading-writing strategy instruction which is easily applicable for teachers.

In the second design phase, together with two history teachers each teaching an 11th-grade class, we co-developed three evaluative writing tasks, to be used as a replacement of regular workbook exercises. Additionally, we developed a strategy on how to perform evaluative writing tasks, which was based on previous studies on reading and writing processes. Our aim was that students would be able to oversee the historical issue at stake, by means of the evaluative prompt, and be able to translate these insights into text. The design process resulted in tangible lesson materials for 11th grade students, in order to improve students' historical writing.

In the third phase, we implemented the constructed design in two groups, and we evaluated its validity, practicality, and effectiveness. We explored effectiveness of the tasks, and the instructions, on students' historical writing, for which we assessed students' texts holistically, on content, structure, and text length.

In the implementation of the design, Principle 1 stood out clearly; reflective interview results indicated that the evaluative writing tasks were perceived as useful learning activities by both students and teachers. Principle 2, the strategy instruction, stood out less clearly. Aspects of the instruction manual were underrepresented or underrated. The strategy elements focusing on *communicating* knowledge, which were so valued from a literacy development perspective, were not fully internalized by the teachers. We consider this our main concern for redesign: to challenge teachers' beliefs system about writing instruction.

With regard to effectiveness, analyses of students' text quality indicated an effect of strategy instruction on *content quality*, however, only in Group 2. With the use of evaluative writing tasks combined with strategy instruction, Group 2 students' content knowledge was better presented in their texts. We found no effects on holistic scores, structure, and text length.

We might question why we found these small effects for Group 2 only. An obvious difference between the two groups was the teacher, however, we found no major differences between the teachers' beliefs system with regard to writing tasks and instruction, or between their interactions with the materials. Moreover, the topics differed; since T2 and T3 dealt with the same topic

(the war in Vietnam), perhaps more history knowledge was required of students in T3 than in T2.

Another explanation might be found in learner variables, which could have influenced the intervention effects on quality. We have conducted the intervention during a challenging educational setting. Although we did not find relations between students' beliefs and quality directly, we did find moderating effects of transactional beliefs. In general, students were demotivated to participate in online class settings. Among students of Group 2 with higher transactional beliefs scores, the motivation was deemed even lower, while we expected these students to be the more persistent ones, as they were conceiving the writing tasks as useful for learning, and thus scoring higher than average.

Furthermore, students' motivation might have been of influence. As Pajares (2003) concluded, students' self-efficacy in writing can influence writing motivation as well as writing outcomes. Our data seemed to confirm this, since we found a direct positive effect of self-efficacy on text quality. However, no *moderating* effect of self-efficacy was found, suggesting students in panel 1 were not motivated any differently than students in panel 2. Another influential factor might have been the amount of writing tasks in the relatively short intervention period of eight weeks. With four writing tasks in two months, students might have felt overloaded. As one student stated in the evaluative interviews: "We were not used to such demanding tasks in history. The tasks took quite a lot of effort" (Student B2).

Next to that, it remains unclear to what extent the combination of the different writing tasks influenced students' historical writing development. From the instructional unit, in which an example task on colonial art was used, we assumed students to transfer their newly obtained reading-writing strategy knowledge to the teacher-designed writing tasks, which were embedded in content lessons. It is possible, however, that all writing tasks, including the pretest, had influence on students' disciplinary writing. Furthermore, students' initial writing proficiency might have been of influence as well, although we did not find such effects in our sample.

To conclude, learning outcomes are promising, yet diffuse. This might be problematic: when a unit is proven effective, this might contribute to changes in teachers' beliefs about addressing disciplinary literacy in their teaching. However, even though the learning outcomes of this study are encouraging, they might not convince teachers completely yet. Therefore, recommendations for redesign are presented in the next section.

6.1 Recommendations for Redesign

Overall, the functionality of the learning materials seemed satisfactory. Both students and teachers were particularly positive about the type of writing tasks. Therefore, recommendations for redesign mainly concern the strategy instruction.

First, from the interviews with students and teachers, we conclude the video needs improvement in pace and variety. The explanatory parts can be shortened, and the video could show more than one student, for example. Also, we would direct teachers to consider to stretch the intervention over a larger time frame, to avoid students feeling overloaded with extensive writing, which might cause motivational problems.

A next recommendation is more fundamental, considering teachers' awareness of design principles and supporting theories, as well as teachers' knowledge. Teachers, in general, are often faced with unexpected events in the classroom, forcing them to adapt their lesson plan to the situation. Teachers must then be able to decide which learning activity can be adapted, and in what way, to still meet the lesson objectives. If the teacher does not have a good understanding of the reasoning behind a certain learning activity, there is a risk of conducting, or adapting, the learning activity in a non-constructive way: it may become much less meaningful. Previous research has shown that individuals rely on heuristics, when situations become complex, and time and information are limited (Westbroek et al., 2020). Thus, a teacher with a clear understanding of his own heuristics, and of the goals behind the material should be the goal. This might be achieved by sharing design principles explicitly. In our study, we mainly informed the teachers about the practical aspects of intervention, and not about our design principles. The teachers were involved in the development of the writing tasks, but not in the development of the strategy instruction. It was remarkable that teachers considered the possibility of shortening, or even deleting, the strategy steps on writing (Steps 5-6-7), since we actually saw these as crucial elements of the intervention. This sense of necessity obviously did not come across, and it clearly showed the knowledge gap we had already reported on in the exploration of the context. Professional development of teachers in these fields thus seems necessary, and has proven its use in previous studies (e.g. Van Drie et al., 2017).

A final recommendation for redesign, is to supply more guidance on how to provide feedback on students' texts. In our study, we left the feedback options open, which led to a focus on content only. When feedback is directed

more towards the writing process itself, this might reinforce the strategy instruction even more.

6.2 The Material's Assets

From our evaluations, we consider two elements of the materials and design process, to be particularly successful: (1) the writing tasks and (2) the co-development of the tasks. The writing tasks used in this study were evaluative in nature, and based on sources. These tasks were considered instructive, and reflecting "the essence of history". Writing thus fitted well with the goals of the subject of history. Furthermore, the tasks were constructed by the teachers themselves. This ensured teacher involvement, and a good fit with the curriculum, which are factors that increase the likelihood of reuse.

6.3 Limitations and Strengths

Although common for a design study, a limitation of the trial study was its relatively small scale, and that it was conducted at only one school. Other limitations regard the circumstances; due to the online educational setting, teachers' supervision while students performed the tasks was not optimal. It resulted, for example, in students writing by hand instead of using the computer. This was not preferred, since previous meta-analyses (Bangert-Drowns, 1993; Goldberg et al., 2003) concluded that texts composed on a computer are generally of greater length, and of higher compositional quality, than hand-written texts. Furthermore, a major disadvantage was that students were not able to write in the classroom, in teacher's sight. This eliminated the general idea behind the inclusion of writing tasks; as supporting students' processes while writing was almost not possible, and students could not collaborate during the pre-writing stage. We thus recommend the trial study to be replicated under more conventional circumstances, with regard to educational practice in the future.

In such a follow-up study, we might want to include measurements for content learning, to obtain more insights in learning outcomes. Furthermore, we might want to obtain more information on students' progress. In our rubric, we consolidated several aspects of content in only one eventual score. This also applied to the concept of structure. However, collecting multiple scores for the separate operationalizations of the constructs "content" and "structure" might be insightful, given that instruction is also focused on these particular aspects.

Despite the small scale of the study, we did find effects on possibly the most important quality measure: content. In the rubric, content quality contained four factors: understanding of the issue, multiperspectivity, argumentation, and use of sources. These factors are important in learning history, and therefore improvements in these factors are obviously imperative for history teachers. Considering the subject's culture of 'content first', this result might be a prerequisite for teachers to actually work on literacy.

Moreover, this encouraging outcome was the result of only a short intervention, consisting of merely three lessons. This was a conscious choice; we aimed to design lessons which were feasible, and had a low threshold for practical implementations. It seems promising that effects can already be achieved with relatively little time investment for history teachers; after all, practicality was one of the design requirements. A potential follow-up study, could be to design a viable and effective professionalization course for teachers, which will not heighten the threshold for implementation of writing instruction into teachers' subject area.

The most important strength of our study might be the set of validated design principles, and intervention materials, for teachers and future educational researchers with a focus on disciplinary literacy. The design principles may offer guidance on how to integrate discipline-specific writing instruction into subject areas. The description of the intervention of Table 1, is a tangible aid for history teachers worldwide, on how writing instruction might be implemented into the history classroom, in only a few lessons. The strategy we developed was based on previous studies on historical reading and writing research, and is usable for students in upper secondary levels. We have attempted to describe the theories behind the principles and learning materials as thoroughly as possible, and we hope to inspire teachers around the world to integrate discipline-specific writing instruction into their teaching.

Apart from these practical implications, this design study also has the potential of strengthening overall knowledge on historical literacy. We have tried to develop a complete, yet compact, discipline-specific Read-Think-Write strategy, building upon previous interventional research. We may contribute to the educational research field with insights into teachers' views on their own role in developing students' historical writing, and on the role of literacy in the history classroom in general.

CHAPTER 4

IMPLEMENTING WRITING PROCESS INSTRUCTION IN THE HISTORY CLASSROOM: A QUASI-EXPERIMENTAL STUDY*

In this study, we have investigated if integrating writing tasks as learning activities within the history classroom, when accompanied by process instruction, would foster students' historical writing proficiency, epistemic experience of writing, and course content knowledge (grades 10-12). We conducted a quasi-experimental study with two experimental conditions ($N = 182$) and a control condition ($N = 86$). Teachers participating in the experimental conditions attended a professional development session to design evaluative writing-to-learn tasks, tailored to their own curriculum. In both experimental conditions, teachers replaced textbook assignment with their new-designed evaluative writing-to-learn tasks. In one experimental condition, teachers added strategy instruction, to support students to perform tasks step-by-step, focusing on disciplinary aspects of inquiry and text. The control condition was a non-writing condition.

Significant effects in a transfer task of the additional strategy instruction were seen in all aspects of text quality, holistically and on discipline-specific criteria, and in epistemic experience. Recall test results showed a similar increase in course content knowledge for the experimental conditions, compared to a non-writing condition. These results imply that writing seems a promising alternative activity for history learning, yet this only results in more advanced writing proficiency when attention is paid to the writing process, by means of strategy instruction.

* Chapter 4 is based on: Holdinga, C. C., Van Drie, J. P., & Rijlaarsdam, G. C. W. (2023). Under-rated yet effective: Writing process instruction in the history classroom. Under review.

This study underscores the potential of reading-thinking-writing process instruction in the content classroom, as a contribution to the development of students' disciplinary literacy skills, without losses for content learning.

1. INTRODUCTION

For every student to become literate – that is one of education's social tasks. In secondary education in the Netherlands, this is an explicit formal goal (Dutch Education Inspectorate, 2022). The idea behind this goal, is that a literate person is more likely to succeed in further education, and in the labour market; moreover, literate individuals are better able to communicate their knowledge. The call for a sound literacy-base has become more urgent now that students' proficiency level seems to decline, according to research of the Dutch Education Inspectorate (2022). Based on these results, the improvement of students' basic skills (reading, writing and mathematics) has been put on the agenda of every school leader.

According to Dutch educational policy, the development of students' literacy is a shared responsibility: all teachers should contribute (Dutch Ministry of Education, Culture, and Science, 2009). However, this is not a shared opinion among teachers. Language teachers are the ones who primarily take responsibility for developing students' literacy, even though language-oriented teaching in subjects has proven its value. From previous studies, it is clear that subject teachers' effort to support students' reading and writing within the disciplines, has additional value to the development of students' literacy (see for example Yore, Bisanz, and Hand's (2003) overview of literacy studies in science). However, in most subjects, the culture of teaching is still primarily focused on subject-specific content knowledge and reasoning. The belief that it is important for a student to be able to communicate content knowledge or reasoning, seems to be underrepresented (Siebert & Draper, 2008).

Biancarosa and Snow argued that "teachers need to realize they are not just teaching content knowledge, but also ways of reading and writing specific to a subject area" (Biancarosa & Snow, 2006, p. 20). To be able to become a subject-area expert, it is necessary to acquire discipline-specific reading and writing heuristics (Monte-Sano et al., 2014; Goldman et al., 2016). And, maybe even more important, it also works the other way around: the development of disciplinary literacy strengthens thinking and reasoning within that discipline. Literacy skills "can act as vehicles for deeper content acquisition" (Miller

et al., 2018, p. 84). Teaching literacy and teaching content do not conflict, they complement and strengthen each other.

Nonetheless there are obstacles. A first obstacle, is that dominant subject area norms often get in the way (Moje, 2008; O'Brien, 1995). The "pedagogy of telling" (see O'Brien et al., 1995) is still dominant in subject teaching; teachers feel pressured to transmit subject content in a timely and effective manner. This results in the view that supporting students in processing source texts and writing texts, is only a time-consuming, additional, burden, in an overloaded instructional agenda (Mac Mahon, 2014).

Teaching literacy is not only felt as a burden, it is also seen as a responsibility for others. Moje (2008) premised that "it is not uncommon to hear teachers in [...] subject areas argue that they should not be expected to assess a student's ability to construct a well-argued essay for their class: 'What matters is the content,' they say, 'I'm not the [language] teacher'." (Moje, 2008: 98). Many teachers initially reject the idea that they are pre-eminently the teachers to highlight the conventions of literacy within their disciplines. Providing writing instruction is thus generally highly underrated by subject teachers.

A second obstacle is knowledge. Teachers are not commonly skilled to teach disciplinary literacy, they feel they have little knowledge about reading and writing processes, and little experience on how to support writing (De Oliveira, 2011). Teachers, although being proficient readers and writers, who are familiar with the academic discourse, may lack awareness of the mental processes needed to interpret texts in their disciplines (Alger, 2007). Consequently, teachers have spent little or no time demonstrating or modeling explicitly how to read and write in their particular subject area. Additionally, preservice teachers are sceptical about the efficacy of teaching disciplinary literacy, offered by disciplinary literacy research. They therefore tend to avoid implementation of such practices (Moje, 2008). However, that is if literacy instruction is even a vital part of preservice teacher training at all, which is often not even the case (De Oliveira, 2011).

In the present study, we aimed to overcome these aforementioned obstacles. What if teachers would make more use of writing as a learning activity in their lessons? And, what if they were supported to provide students with writing process support? In the design study of Chapter 3, we have developed a discipline-specific writing strategy instruction. In the current study, we added a professional development program aimed at teaching how to design writing tasks, to implement writing instruction, and to support students' discipline-specific writing process, with the overall aim of improving students'

disciplinary source-based writing. We focused on students in pre-university level, upper secondary school (grades 10-12, age 15-18), as upper secondary school levels provide opportunities to work with more extensive writing tasks concerning more complex issues, aimed at profound understanding of content. This way, reading, writing, and content learning, would be intertwined. Although the obstacles mentioned might seem surmountable issues, they are deeply rooted in teachers' and students' routines. Therefore, in the current study, we tried to adapt to the existing classroom routines in two ways. First of all, teachers developed writing tasks tailored to their own curriculum. This ensured ecological validity. We considered this a strength of the study; if the intervention turned out to be effective, it might be in other situations as well, with other teachers, in other schools. Secondly, following Moje (2008), in our study we have tried to foreground the discipline. In our case, this was the subject of history, which is known for its persistent subject culture of "content first". As Gillis (2004) advised, we did not adopt generic reading and writing strategies into the history classroom, but we adapted reading and writing strategies into a discipline-specific strategy for historical writing based on historical sources. Source-based writing is a common task in history classes in upper secondary grades, yet linguistically demanding for students.

1.1 Linguistic Challenges in History

The school-subject history comes with many linguistic challenges for students, concerning both reading and writing. First of all, textbooks contain multiple text types: informative texts which are full of dense and abstract language, and source texts, especially primary sources, which are often difficult to read, and open to interpretation. Students are expected to read and assess historical sources of different kinds: these might be texts, but can also be cartoons, photos, or graphics. Sourcing is an important skill to develop in secondary school. When multiple documents are involved, it is especially important for students to attempt to contextualize the information from each document, to connect information elements from different texts, and to create a coherent, integrated model necessary for understanding (Wiley et al., 2014). Furthermore, students must be able to discuss complex issues in oral or written form, based on source inquiry. Although several text genres common to history can be distinguished (Coffin, 2006), the most prominent is argumentative writing: "Argumentative historical writing, through which historians defend their interpretations, their use of evidence, their research methodologies, and the significance of their work, represents the pinnacle of historical

writing" (Nokes & De La Paz, 2018, p. 559). Reading and writing thus play a significant role in the learning of history, however, students experience these to be challenging (Schleppegrell & De Oliveira, 2006; Goldman et al., 2016).

Although writing is not grounded as a learning activity within the history classroom in the Netherlands, it has proven its great potential for learning in general (Graham et al., 2020), as well as for learning history specifically (Voss & Wiley, 1997; Wiley & Voss, 1996, 1999). Previous research of Monte-Sano (2008) has shown that simply adding writing into content lessons is not sufficient for growth in historical writing. Strategy-instruction on reading strategies (Nokes et al., 2007; Reisman, 2012; Van Driel et al., 2022b), and reading-writing strategies (De La Paz, 2005; De La Paz & Felton, 2010; De La Paz et al., 2014) have proven to be effective for the development of historical reasoning. Therefore, the aim of the present study was to improve students' writing proficiency through (1) inclusion of source-based writing as a learning activity, and (2) a discipline-specific writing strategy intervention in the history classroom, considering the obstacles encountered in practice.

1.2 Research Questions

Our research questions were:

RQ1: To what extent does the replacement of workbook assignments by evaluative writing tasks contribute to (a) disciplinary source-based writing proficiency, (b) epistemic experience of writing, and (c) students' course content knowledge?

RQ2: To what extent does the addition of strategy instruction to such writing tasks contribute to (a) disciplinary source-based writing proficiency, (b) epistemic experience of writing, and (c) students' course content knowledge?

We hypothesized that the replacement of workbook assignments by evaluative writing tasks, as knowledge constructing learning activities (RQ1), would not result in higher-quality texts (Monte-Sano, 2008), but that that it might result in more epistemic experience, a term derived from Bereiter (1980). With the epistemic experience of writing, we refer to the awareness of acquiring and constructing knowledge through writing. We hypothesized that students who would perform evaluative writing tasks as learning activities, would experience the epistemic function of writing. Therefore, also in transfer tasks, we expected that students would experience writing as a learning activity,

resulting in a greater epistemic experience, more than students in a non-writing condition.

Furthermore, we predicted that students would gain more content knowledge of the topics they wrote about, by means of more extensive writing tasks, than by other (non-writing) learning activities, since the act of writing has the potential of developing a more deep and complex knowledge (Graham et al., 2020).

Regarding RQ2, we expected students who received strategy instruction to write higher quality texts (Graham & Perin, 2007), than students who only performed writing tasks without strategy instruction, or students in a non-writing condition. We also expected the strategy instruction to reinforce effects on epistemic experience, since the strategy instruction would ease the writing process, and therefore create more space for content development. Hypothetically, for students who received strategy instruction, this would occur both in transfer tasks, resulting in greater epistemic experience, as well as in the writing tasks used as learning activities, resulting in more course content knowledge.

To examine whether effects of conditions interacted with learner characteristics, we added a third, explorative research question:

RQ3: To what extent are effects on RQ1 and 2 moderated by learner characteristics (writing beliefs and self-efficacy)?

We aimed our lesson materials to be equally effective for all students, regardless of belief types or self-efficacy levels with regard to writing. However, learner variables could moderate intervention effects. Previous research on writing beliefs from White and Bruning (2005) has defined two main belief types for writing: transmissional beliefs, which are associated with the perception that writing is mainly a means of transmitting information to one another; and transactional beliefs, which are associated with the conception that writing might fulfil an epistemic function. Spanish researchers concluded that students with high transactional and low transmissional writing conceptions, might be expected to write higher quality texts (Villalón et al., 2015). We thus hypothesized that students with high transactional beliefs would write higher quality texts on all measurement occasions, and would experience writing more often as an epistemic experience.

A second learner variable, which might have direct effects, and also moderate the effects, is students' self-efficacy. Previous research has shown self-

efficacy to be positively associated with writing performance (Pajares, 2003). In our study we aimed to explore the possible moderations of self-efficacy, by assessing if, for example, students with low self-efficacy for writing would benefit more from the strategy instruction, since they potentially require more hold for the writing process, or if this works inversely.

2. METHOD

2.1 Research Design

We implemented a quasi-experimental pre-test/post-test design, with three conditions: two experimental conditions: Writing Task and Strategy (WT+S), and Writing Task (WT), and a Control Condition (C). In the WT+S Condition, the teachers attended a professional development (PD) session about epistemic writing task-design. In order to replace textbook assignments, they designed such writing tasks themselves, under researcher's guidance and monitoring, and implemented a writing strategy instruction. In the WT Condition, the teachers also attended the same PD-session, but they were not informed about the strategy instruction, and did not implement it. In the Control Condition, the teachers provided regular history education, without evaluative writing tasks or strategy instruction. In all conditions, the teachers taught the history content as pre-planned in the school specific year plan, which varied between groups. We considered this a strength of our design; it enhanced ecological validity, and generalizability across thematic contents. An overview of the research design is presented in Table 4.1.

2.2 Participants

In this study, 11 qualified history teachers, from 10 different schools, participated with 14 groups. Initially, 12 teachers participated in this study, with 16 groups, however, in two groups taught by the same teacher, data collection procedures were not executed according to established criteria. Therefore, these two groups were excluded from further data collection and subsequent analysis.

Since motivation of the teacher was considered an important factor for successful educational innovation, teachers were offered a choice with regard to their preferred experimental condition, i.e, WT+S or WT. Subsequently, other teachers were engaged to participate in the Control Condition; these teachers received the experimental lesson materials following the post-test, for them to potentially implement in their subsequent lessons, if preferred.

Table 4.1. Research Design

Condition	N = 268	PD session	Pretest	Intervention		Posttest
				Writing tasks	Strategy instruction	
WT+S	n = 119 7 groups 6 teachers	+	+	+	+	+
WT	n = 63 3 groups 2 teachers	+	+	+		+
Control	n = 86 4 groups 3 teachers		+			+

Note: When an activity was conducted, a + is presented.

The students participating in this study were enrolled in the pre-university program (grades 10-12). Group-size differed from 10 to 32 students ($\mu = 20.6$). Due to Covid-19 restrictions (quarantines or isolations), the rate of students' absence was higher than anticipated, and, due to this, insufficient data was available for 10 students. All students – and in the case of minors, also their parents – actively consented for participation. When consent was withheld (by either students or parents), students did participate in the classroom activities during the intervention period, since these replaced regular activities, but their data were not collected. Overall, from ten students (WT+S: 2; WT: 4; C: 4) no consent was obtained; there were no parents who objected to participation of their child.

The results presented in this study were based on the data of 268 students (14-19 years, $\mu = 16.2$, $sd = .85$).

2.3 Procedures

2.3.1 Professional Development Session

The nine teachers who taught the experimental conditions (WT+S and WT), took part in an online one-hour PD session, about teaching writing in history, and specifically, writing task-design. In this session, two researchers explained and discussed the background of the study, the main ideas behind the lesson materials, and four, hereafter described, writing task design principles.

Principle 1: it is advised to make use of evaluative questions, which ask for inquiry of different perspectives on an issue. Such tasks have the potential to develop understanding of content (Newell & Winograd,

1995; Voss & Wiley, 1997) and to elicit historical reasoning (Monte-Sano & De La Paz, 2012; Van Drie et al., 2006).

Principle 2: writing from sources is advised, since reading and writing may reinforce each other (Graham et al., 2018; Wiley & Voss, 1996). For historical writing, it is best to preselect, preferably, primary sources, representing multiple perspectives on the issue presented in the evaluative question (Britt & Rouet, 2012).

Principle 3: the writing task must replace regular workbook assignments, and are not 'extra'. This way the writing task becomes an actual writing-to-learn task, instead of a simpler processing, or reproduction task, of smaller content units.

Principle 4: the writing task can be completed within one lesson hour. Rationale behind this principle was that students' writing processes become visible for the teacher, since the writing would take place in the classroom.

The teachers themselves designed the writing tasks based on these four principles, to use during the intervention period. They tailored the tasks to their own school-specific year plans, with regard to the content of the tasks. The research team supported the teachers by providing written or oral feedback on draft versions. Generally, it was advised that each teacher developed two writing tasks for their units. An overview of topics and evaluative questions, is presented in Table 4.2. To assess the extent to which the developed tasks complied with the design principles, all tasks were coded (Table 4.3). As presented, teacher C developed only one writing task, which was more extensive, and was without preselected sources, because they felt it suited their students' grade (12th) better to let each search for appropriate sources themselves. It was decided that all developed tasks, including the task of teacher C, were of high enough quality to be included in further analyses.

Table 4.2. Overview of Topics and Tasks

Condition	Teacher	Group	Grade	Topic during intervention period	Evaluative question (Task ID)
WT+S	1&2	1&2	11 th	Modern imperialism related to industrialization	To what extent did Industrialization lead to a higher standard of living? (A1) To what extent was the ethical politics experienced as a continuation of the Dutch domination of the Dutch East Indies? (A2)
	3	3&4	10 th	The expansion of religions in the Middle Ages	How important was the battle of Poitiers in halting the advance of the Saracens? (B1) To what extent were crusaders guided by their own interests when participating in the crusades?(B2)
	4	5	12 th	Decolonization	Open assignment: students formulated their own inquiry question and searched for suiting sources themselves. (C)
	5	6	11 th	The conflict in the Netherlands that resulted in the foundation of a Dutch state	To what extent was the relationship between Philip the Good and Ghent's citizens around 1450 illustrative for the relationship between princes and city councils in the same period? (D1) To what extent was the Iconoclasm of 1566 an expression of religious, political and/or socio-economic discontent? (D2)
	6	7	11 th	The Age of Enlightenment	To what extent is it justified to introduce compulsory vaccination in a pandemic, starting from ideas of John Stuart Mill? (E1) May a Prime Minister call upon citizens to have their children vaccinated against COVID, if you take as your starting point the ideas of the Enlightenment philosophers Locke and Rousseau? (E2)
	WT	8	9&10	11 th	The division of the world into two ideological blocs
9		11	11 th	The Soviet Union	Which of the causes of World War I do you think was most important? (G1) To what extent was the Soviet Union a workers' paradise in the period 1924-1941? (G2)
C	10	13	11 th	The Treaty of Versailles	-
	11	14	10 th	Crusades	-
	12	15	11 th	The time of regents and monarchs	-
		16	10 th	The time of monks and knights	-

Table 4.3. Coding Scheme for Teacher-Developed Tasks

Design Principles	Task ID													
	A1	A2	B1	B2	C	D1	D2	E1	E2	F1	F2	G1	G2	
<i>#1 Evaluative question</i>														
Present yes/no	yes	yes	yes	yes	yes*	yes	yes	yes	yes	yes	yes	yes	yes	yes
<i>#2 Multiple sources</i>														
Present yes/no	yes	yes	yes	yes	yes*	yes	yes	yes	yes	yes	yes	yes	yes	yes
Number of sources	6	7	7	5		5	6	2	4	5	5	4	5	
Number of textual sources	3	5	7	5		4	6	2	3	4	4	2	3	
Number of visual sources ¹	3	2	0	0		1	0	0	1	1	1	2	2	
Mean length of source texts	201	183	211	168		200	162	455	312	175	146	131	105	
<i>#3 Fit</i>														
Students' grade level	11 th	11 th	10 th	10 th	12 th	11 th	11 th	11 th	11 th	11 th	11 th	11 th	11 th	
Difficulty as experienced by teachers ²	3	2	2	2	2	2	2	2	2	2	2	2	2	
<i>#4 Task duration</i>														
Appropriate for one lesson hour ³	2	2	3	3	3	2	2	1	1	2	2	2	2	
Duration of lessons	50'	50'	45'	45'	60'	50'	50'	50'	50'	50'	50'	40'	40'	

¹ Pictures, tables, cartoons, or videos.

² Difficulty: 1 = easy, 2 = not too difficult, not too easy, 3 = difficult.

³ Task duration: 1 = too short for one lesson hour, 2 = not too short, not too long, 3 = too long for one lesson hour.

* Students formulated their own evaluative question and searched for appropriate source materials themselves.

2.3.2 *Description of the Intervention*

In Condition WT+S, strategy-instruction as well as evaluative writing tasks were implemented. In Condition WT, such evaluative writing tasks were implemented without strategy instruction. The interventions as designed for Condition WT+S and WT are presented in Table 4.4.

The strategy instruction for Condition WT+S consisted of two lessons. The instructional design (how to teach the strategy) was roughly based on Harris and Graham's classic SRSD-framework (1996), and consisted of five instructional stages: (1) development of background knowledge, (2) describe it, (3) model it, (4) support it, and (5) independent performance.

In Lesson 1, direct strategy instruction was provided. Central to this was a history-specific strategy for performing evaluative writing tasks: the Read-Think-Write strategy (see Chapter 3). This RTW strategy, is a seven-step strategy directing students how to read source texts diligently, how to connect information from different sources, and how to write a text with the reader in mind. In the second instructional stage, model it, it was advised to show students how a strategy is used. In an attempt to unburden the teachers - who might feel challenged by having to model unfamiliar reading and writing processes - a video with students modeling the RTW strategy was provided for usage. Each teacher could decide to simply show the video on the smart board, and discuss it in a class discussion.

In Lesson 2, students read three example texts, and discussed them with peers. In a class discussion, a list of criteria for good evaluative history texts was constructed.

After the two strategy lessons, students from Condition WT+S performed the writing tasks as developed by their own teacher. Each task was discussed in class in a feedback session, with the feedback focusing on the writing process. In Condition WT, students performed two writing tasks, as developed by their own teacher, and these writing tasks were discussed in feedback lessons with product feedback, focused on content. For both conditions, students were allowed to collaborate in the prewriting phase of Lesson 3 and 5 (Table 4.4), but they wrote their texts individually. Writing tasks were all performed on a computer, using laptops in the classroom, or computers in a computer room.

Table 4.4. Description of the Intervention

Lesson	Stage	Learning activities	Condition WT+S	WT
0	<i>Develop background knowledge</i>	Task experience	Students perform a first evaluative writing task on an uninstructed historical issue (= pre-test).	A
1		Reflection	The students write down what they thought was easy while performing the task from Lesson 0, and what was difficult.	-
	<i>Describe it</i>	Building new knowledge	The teacher presents and explains the <i>Read-Think-Write</i> strategy.	
		Comparing strategy to own experience	The students compare their own experiences with evaluative tasks to the presented strategy.	
	<i>Model it (process)</i>	Observing strategy demonstrated by a modeling peer (video)	Students watch a 12-minute video in a plenary session on the main screen. This video presents the strategy and contains fragments showing modeling peers, who demonstrate how each step of the strategy could be performed.	
		Assessing a peer's performance	As a processing activity, the students individually compare the performance of the modeling peer with the strategy as presented by scoring the peer on a scale from 0 to 100.	
2	<i>Model it (product)</i>	Assessing peers' texts	Students individually assess three example texts.	-
		Generating criteria	In a class discussion, students generate a criteria list.	
		Applying new learning	Students apply the criteria; they revise the text they wrote (Lesson 0), with the criteria list in mind.	
3	<i>Support it</i>	Scaffolded practice	Students perform a next evaluative writing task, scaffolded by the support of peers and the teacher.	A
			Written guide available.	B
4	<i>Support it</i>	Teacher/Peer feedback	The teacher chooses how to provide feedback from the suggestions in the teacher manual.	A
			Feedback is focused: on <i>process</i> .	C
5	<i>Independent performance</i>	Individual work with the help of a written guide	Students perform a third evaluative task.	A
			Written guide available.	B
6	<i>Support it</i>	Teacher/Peer feedback	The teacher chooses how to provide feedback from the suggestions in the teacher manual.	A
			Feedback is focused: on <i>process</i> .	C
7	<i>Transfer</i>	Individual work	Students perform a fourth evaluative source-based writing task on an uninstructed historical issue (post-test).	A

Note: A = As WT+S; B = No guide; C = Feedback is focused on content.

As presented in Table 4.4, Lessons 4 and 6 centered on feedback on students' texts. The teachers were provided with condition specific suggestions on how to provide this feedback. These suggestions can be found in the supplementary materials. The feedback lessons were designed to be low cost for teachers, since it is generally known that the workload of providing feedback on written products is one of the main reasons for teachers to avoid writing tasks as learning activities. In the teaching manual for Condition WT+S, the suggestions all encouraged the teachers to focus their feedback on students' writing process. In the teaching manual for Condition WT, teachers were guided towards providing product feedback focused on content.

2.3.3 Control Condition

Between the pre- and the post-test, students in the Control Condition attended regular history classes, as they were used to. This mostly involved the teacher explaining topics within the classroom, and students working on assignments from the workbook, individually or in groups. Students wrote notes and short answers to questions, but no larger writing assignments were performed.

2.3.4 Implementation Fidelity

To check whether the intervention was implemented as intended, we included triangulated measures: the research team collected teacher logs, did observations, and conducted reflective interviews with both students and teachers. Teachers from all three conditions kept a guided logbook during the intervention period. In Conditions WT+S and WT, LH observed one intervention lesson for each group. After the intervention period, LH interviewed each teacher of Condition WT+S and WT, as well as a group of three to four students from each group within these conditions. Interview guides can be found in Appendix E.

The teachers were free to plan their lessons as they preferred, resulting in different durations of the intervention. The mean duration between pre- and post-test was 8.3 weeks ($sd = 3.2$ weeks). A consequence of the free planning was that the number of history content lessons also differed between groups and conditions.

Concerning the intervention lessons, the reflective interviews and observations exposed several deviations from the intervention-as-designed in Groups 5 and 7. In Group 5, there was only one extensive writing task instead of two (as previously discussed). Since this task was more time-consuming, the order of activities was adjusted according to this. The task was explained

at the beginning of the intervention, as students needed time to think of an evaluative question, and to search for source materials. The students wrote their texts after Lesson 1, they then discussed example texts in Lesson 2, and students were given the opportunity to revise their respective texts based on this discussion, prior to handing in their work for a formal assessment. The students all received individual feedback, as the task was a formal exam, and thus there were no feedback lessons.

Group 7 deviated from the design as constructed, due to presumed lack of time: the students performed the two evaluative tasks (Lesson 3 and 5) both in one 50-minute session. Both tasks were discussed in the subsequent lesson in one session as well, thus, Lesson 4 and 6 were also merged.

Furthermore, Groups 8 and 12 were not provided with sufficient time for the pre-test. Their intervention-as-implemented was also different from intended; the students worked individually, and on their own pace, through the materials. Therefore, these two groups were excluded from further analyses.

Regarding the options for providing feedback, the options each teacher chose for the two feedback lessons were coded (Table 4.5). Next to the five suggestions presented in the teaching manual (coded #1-#5), there were two more codes. Code #6 resembled personal feedback. Some teachers chose to provide students with personalized written feedback, despite our discouragement, due to lack of lesson time to discuss example texts during lesson hours. In some cases, the teachers graded students' texts, and therefore chose to provide written feedback as well. Code #7 resembled only a general discussion of content. For instance, some teachers discussed which perspectives should be mentioned in the text, in order to discuss the issue properly. The teachers who gave feedback coded as #7, explained this was due to time constraints.

As can be appreciated in Table 4.5, three teachers chose to discuss mainly content after the second writing task. However, they had already discussed the first task extensively.

Thus, all teachers except for Teacher 7 (Groups 8 and 12) implemented Lesson 1 to 5 of the intervention as intended, as all essential elements were implemented, and this implementation was of sufficient quality. Lesson 6, however, was not considered optimal by three of the teachers. These teachers explained that lack of time played an important role. However, when a teacher is confronted with time constraints, this means they are forced to prioritize. Priorities were apparently not set on providing feedback on students' texts, but more on knowledge accretion. For example, teacher 9 explained that they had read students' texts for task F2, and then decided to skip lesson 5:

Teacher 9: "I think I was shocked at how they were lacking in content in their texts, and then I thought: I have to focus on that now, because otherwise their final test results will be a complete disaster."

Nevertheless, it was decided to include the results of these three teachers as well. Feedback was of lesser quality after the second task, but not absent in their intervention as a whole.

Table 4.5. Coding of Feedback Lessons

Condition	Teacher	Group	Feedback Suggestion				
			1	4	5	6	7
WT+S	1	1			A1	A2*	
	2	2		A1		A2*	
	3	3&4	B1			B2	
	4	5				C*	
	5	6	D1				D2
	6	7	E1&E2				
WT	8	9&10	F1				F2
	9	11	G1				G2

Notes: Feedback suggestions: (1) creating a scale; (2) comparing text to a scale (not used); (3) discussion of random text (not used); (4) discussion of good example; (5) assessing on criteria; (6) personal feedback; (7) discussion with focus on content only. For the task-ID's, see Table 4.2.

*Texts were graded by the teacher.

2.4 Variables and Measurements

To answer the research questions, the effects of the intervention on three aspects were measured. Firstly, the students' historical writing proficiency (RQ1-2a) was measured, using source-based writing tasks (pre-test/post-test). Secondly, the epistemic experience (RQ1-2b) was measured, using a statement to rate about the epistemic experience immediately after performance the writing task (pre-test/post-test). Thirdly, course content knowledge (RQ1-2c) was measured, using a recall test (pre-test/post-test). Furthermore, to answer RQ3, all students completed two questionnaires before the start of the intervention.

2.4.1 Writing Proficiency

To measure source-based writing proficiency (RQ1-2a), students performed writing tasks about history topics which were unrelated to course-content. The post-task might thus be considered a transfer task. In the pre-test, the main question was: To what extent do you think Western museums should

return colonial art to the country of origin? Four source texts were provided ($\mu = 218$ words). For the post-test, the task was about the Dutch police actions in Indonesia (1947-1949)*. The main question here was: To what extent can the term "police action" be justified, in hindsight? Again, four source texts were provided ($\mu = 198$ words). Both tasks contained mostly primary sources, representing multiple perspectives. Each task took about 30 minutes to perform.

We controlled for condition effects on task perception on four dimensions: prior knowledge, topic interest, effort and difficulty. This was done by asking all students to rate four statements (Table 4.6), on a scale of 1 (completely disagree) to 5 (completely agree), in order to explore how each student had perceived their tasks.

Table 4.6. Task Perception Measurement

Prior knowledge	<i>I knew a lot about the topic already before I'd read the source texts.</i>
Topic interest	<i>I think the topic of this task is interesting</i>
Effort	<i>I put a lot of effort into this writing task.</i>
Difficulty	<i>I thought this writing task was difficult.</i>

2.4.2 Epistemic Experience

A measurement to obtain insight into the students' knowledge gain of the writing task topic, was incorporated into a question at the very end of the pre-/post-test on epistemic experience (RQ1-2b). We asked students to rate the statement "Through this writing task I learned a lot about the topic" on a scale of 1-5. "The topic" referred to the topic of the respective source-based writing task. We asked this question to explore if the students felt they had gained knowledge on a topic, unrelated to the lesson content, through the task set.

2.4.3 Course Content Knowledge

To measure the effect of the replacement of textbook tasks by writing tasks, and the addition of strategy instruction, on course content knowledge (RQ1-2c), we administered a ten-minute open recall test, as the first measurement in the pre-test and the post-test sessions (Casado-Ledesma, et al., 2020; Langer & Applebee, 1984). The question used was: "What do you know about ...". On the dots, each teacher filled out the topic they discussed in the upcoming (pre-

*In 1947, the Dutch government did not recognize the Republic of Indonesia as an independent state, but regarded it as a rebellious movement within the colony of the Dutch East Indies, against which "police action" had to be taken.

test) or past (post-test) history lessons, during the intervention period. These topics differed per group (Table 4.1).

2.4.4 Learner Variables

We administered two questionnaires before the pre-test session: a writing-beliefs questionnaire, and a self-efficacy questionnaire (RQ3). The teachers instructed their students to complete these in a lesson hour prior to the start of the intervention. It took students about 10 minutes to complete both questionnaires, using Qualtrics. Both questionnaires can be found in Appendix C.

The writing-beliefs questionnaire consisted of 26 statements about writing in general. We distinguished four aspects (transmission; high amount of revision; emotional engagement; cognitive engagement), based on the validated questionnaire of Vandermeulen (2020). All students were asked to indicate to what extent they thought each statement was true, on a scale from 1 (totally disagree) to 5 (totally agree).

The self-efficacy questionnaire was developed and tested in a former study (Chapter 3). It consisted of 30 'I can'-statements, which were all related to aspects of the historical reading-writing process. All students were asked to indicate, to what extent they thought the statements were true, on a scale from 0 to 100. The consistency of the questionnaire was high (Cronbach's alpha .93). Although a factor analysis uncovered the structure of six factors (for details, see Appendix F), we maintained students' total self-efficacy scores in our analyses, based on our exploratory goal.

2.5 Data Preparation

2.5.1 Writing Proficiency: Text Rating Procedures

Students' texts were assessed in two ways: analytically, using a rubric with five dimensions, and holistically, using a scale with annotated anchor texts. With the rubric, assessors were asked to score the students' texts on five dimensions: situational understanding, multiperspectivity, argumentation, source use, and structure (derived from De La Paz et al., 2017). Subsequently, they scored each text holistically, with the help of a scale. This scale was developed by the research team for a previous study (presented in Chapter 3), and contained five anchor texts, representing scores 70-85-100-115-130. The anchor texts in the scale were accompanied by annotations, pointing towards five foci: the dimensions of the rubric. It resulted in six scores for each text. Lastly, a final indicator for text quality was text length (number of words).

The data set consisted of 277 texts. Per topic, the texts were randomly divided over sets containing 8-10 texts. We excluded 11 texts (pre-test: 5; post-test: 6) which were unfinished (containing only a few words accompanied by a remark such as 'I had no time left'), or were a reflection of the task not being fully understood (for example when students noted: 'I have no idea what to do with this task' or 'I don't understand the source texts'). Of these excluded texts, nine were from the control condition group.

Because the data set consisted of a large number of texts, the assessments were performed by three jury panels. In total, 21 history teachers (e.g., former teachers, teacher trainers, teacher students) as well as LH, took position in the jury panels. A system was set-up, in order to have each jury member rate six sets (= approximately 60 texts in total, 30 texts from the pre-test, 30 texts from the post-test), while (partly) overlapping with four other raters. This way, each text was assessed by three jury members. Intraclass correlation coefficients of the juries were between .79 (pre-test argumentation) and .89 (post-test structure) ($\mu = .84$), which was considered satisfactory, and further analyses were consistently conducted with the mean of the three jury's scores.

2.4.2 Course Content Knowledge: Recall Test Analyses

It was hypothesized that students with more knowledge-gain, would write more new information during the post-test. Therefore, each student's pre- and post-recall tests were compared, and the new T-units (minimal terminable units (Hunt, 1965)) were counted. For example, if a student wrote down three T-units in the pre-test and three in the post-test, the number of T-units from the post-test that contained new or extended information compared to the pre-test were recorded. When there were three new information units in the post-test, the student was given the score '3'. T-units had to be correct, and relevant, to be counted as 'new'. A subset (10%) was analyzed by a second rater, in order to check the reliability of scoring (ICC = .95).

Secondly, the mean length of new T-units was calculated, since it was hypothesized that students in the writing conditions would write richer, and more new units containing connectors and adverbs, indicating more relations between concepts. Generally, relations within units are associated with more elaborated knowledge, and understanding of content.

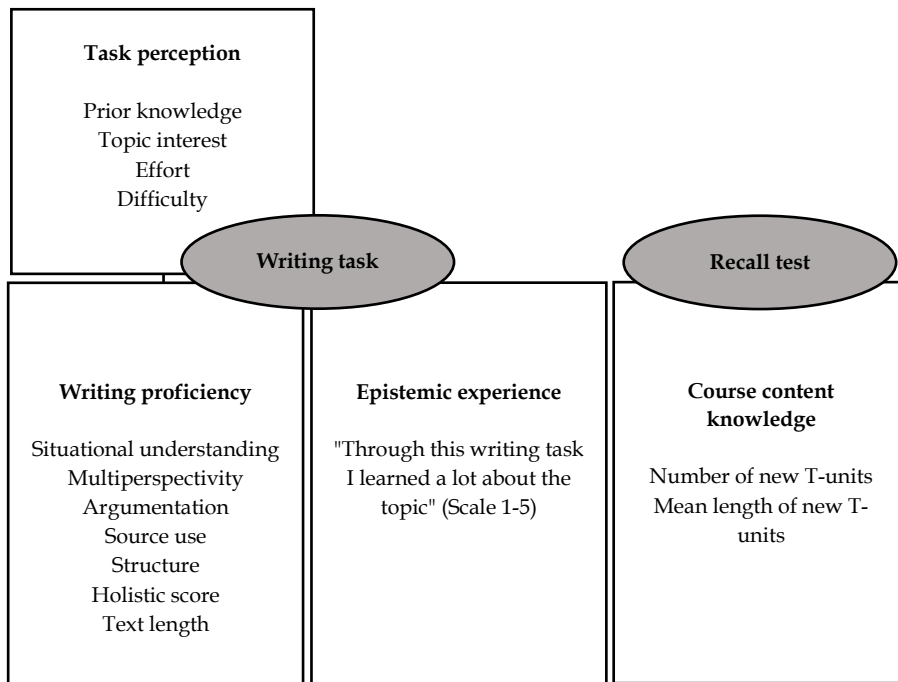
Those students' recall tests which were completely blank (i.e., 2.3%) were excluded from the analyses, as the underlying reasons were unknown. Conversely, responses only consisting of remarks, such as 'I don't know anything

about this', were included: wit the word/T-unit count for tests only containing such a remark being 0.

2.5 Data Analysis

We conducted a multilevel analysis for three dependent variables: (1) writing proficiency, (2) epistemic experience, and (3) course content knowledge. An overview of variables is presented in Figure 4.1. The seven indicators for text quality were closely related (r between .77 and .90). The holistic score correlated best with the other six indicators (r between .83 and .90), based upon it was decided to focus the analyses on this score, when analysing writing proficiency.

Figure 4.1. Overview of Variables



The effects of the interventions were tested by comparing four nested models, with Group as random factor, and time as repeated measure, including Group*Subject as random factor: Model 0 with random factors only, Model 1 with the added factor Time, Model 2 with the added factor Condition, and Model 3 with the added interaction Time*Condition. To gain insight into the effects of the Condition at the post-test measurement, the analysis for the best fitting model was reran, using dummy variables for the effect of Condition (three levels), and Time (two levels), and the resulting pair-wise comparisons were checked.

For the course-content knowledge indicators (number of new T-units and length of new T-units), two models were compared, while Time was not a factor: Model 0 with Subject as random factor, Model 1 with added factor Condition.

To explore moderating effects of self-efficacy and beliefs, we centered the scores, and expanded the models presented above, with Model 4 (general effect of the learner variable LV); Model 5 (interaction LV with Time); Model 6 (interaction LV with Condition); and Model 7 (interaction LV with Time and Condition). We ran these series of models for holistic score (for writing proficiency), and epistemic experience. All model comparisons can be found in Appendix G.

2.6 Preliminary Analyses

2.6.1 Pre-test Differences

We conducted preliminary analyses, to explore differences between conditions, on all outcome variables (all writing proficiency variables, epistemic experience, and all three variables for course content knowledge), at the pre-test. For none of the dependent variables, the learner variables, and task perception variables, effects of condition were observed at the pre-test.

2.6.2 Task Perception

Because the pre-test task was an essential part of the intervention lesson (Lesson 0), we could not randomly assign pre- and post-test topics. Topics were thus nested within measurement occasions. Therefore, we checked for pre-test/post-test similarity, on four aspects: prior knowledge, topic interest, effort, and difficulty. Mean scores for the four statements on task perception, are presented in Table 4.7.

A multilevel analysis revealed effects of the pre-test/post-test writing task. We found differences on two out of four aspects. Students felt they had more

prior knowledge in the post-test than in the pre-test ($F(1,223) = 9.73, p = .002$), and they perceived the post-test as more difficult ($F(1,235) = 22.07, p < .001$). Additionally, the students perceived the two tests as similarly interesting, with a relatively high mean score, indicating the tasks were indeed experienced as quite interesting. The students also indicated that they had put similar effort into their performances of the either of the two tasks. We found no interaction effects with condition, meaning that possible differences between conditions on outcome variables were not influenced by differences in these task-perception variables.

Table 4.7. Task Perception Scores (Scale 1-5)

	Condition WT+S		Condition WT		Control Condition	
	Pretest <i>M (SD)</i> <i>n</i> = 107	Posttest <i>M (SD)</i> <i>n</i> = 91	Pretest <i>M (SD)</i> <i>n</i> = 55	Posttest <i>M (SD)</i> <i>n</i> = 55	Pretest <i>M (SD)</i> <i>n</i> = 70	Posttest <i>M (SD)</i> <i>n</i> = 78
Prior knowledge	2.58 (.98)	2.71 (1.04)	2.78 (.92)	3.02 (1.05)	2.39 (.90)	2.80 (1.10)
Topic interest	3.54 (.92)	3.64 (.89)	3.67 (.90)	3.82 (.82)	3.44 (.92)	3.51 (1.00)
Effort	3.38 (.79)	3.45 (.73)	3.19 (.58)	3.16 (.71)	3.07 (.78)	3.30 (.94)
Difficulty	2.74 (.91)	3.00 (.90)	2.59 (.84)	3.06 (.91)	2.71 (.99)	3.14 (1.04)

2.6.3 Learner Variables

The students' mean scores for learner variables are presented in Table 4.8. For self-efficacy, the mean individual total score was 66.2 ($N = 242$, $\min = 39.7$, $\max = 95.5$, $sd = 10$). For beliefs, the scale "revision" was not reliable ($\alpha = .47$), similarly to in the studies of Vandermeulen (2020). Based on this, no further analyses with this scale were conducted. The three remaining beliefs scales did not correlate highly (varying from $-.031$ to $.314$), which lead us to consider them separate. There were no significant differences between conditions.

Table 4.8. Mean Scores for Beliefs (Scale 1-5) and Self-Efficacy (Scale 0-100)

	Condition WT+S <i>n</i> = 115 <i>M (SD)</i>	Condition WT <i>n</i> = 61 <i>M (SD)</i>	Control Condition <i>n</i> = 78 <i>M (SD)</i>
Transmission (7 items, $\alpha = .67$)	2.88 (.56)	2.92 (.52)	2.90 (.51)
Emotional Engagement (5 items, $\alpha = .59$)	3.36 (.53)	3.35 (.56)	3.33 (.48)
Cognitive Engagement (3 items, $\alpha = .77$)	3.65 (.77)	3.80 (.63)	3.55 (.80)
Self-efficacy	65.6 (9.5)	68.9 (10.6)	65.2 (10.1)

3. RESULTS

In the following section the intervention effects on students' writing proficiency, epistemic experience, and course content knowledge are presented. The focus lays on the differences between conditions, as pre-test differences were checked for with preliminary analyses.

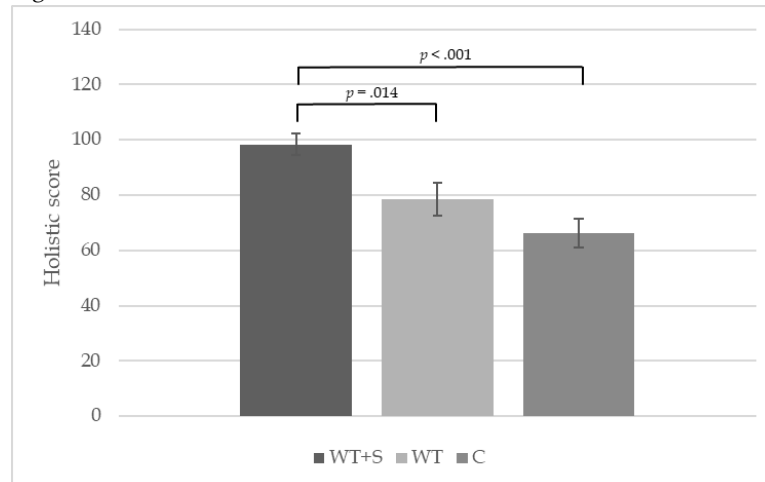
3.1 Writing Proficiency (RQ1-2a)

For holistic score, Model 3 fitted the data best, as no differences were observed at the pre-test. This revealed an interaction effect of Time and Condition, and therefore an effect of conditions on the post-test (Figure 4.2).

In the post-test, the WT+S Condition scored significantly higher than the WT Condition (mean difference: 19.68, $se = 7.1$, $p = .014$), and than the Control-condition (mean difference: 32.02, $se = 6.47$, $p < .001$). Condition WT did not differ significantly from the Control Condition with regard to holistic scores ($p > .05$).

Moreover, similar results were observed for all other indicators of text quality: situational understanding, multiperspectivity, argumentation, source use, structure, and text length.

Figure 4.2. Holistic Scores in the Post-test, as Estimated Under Model 3

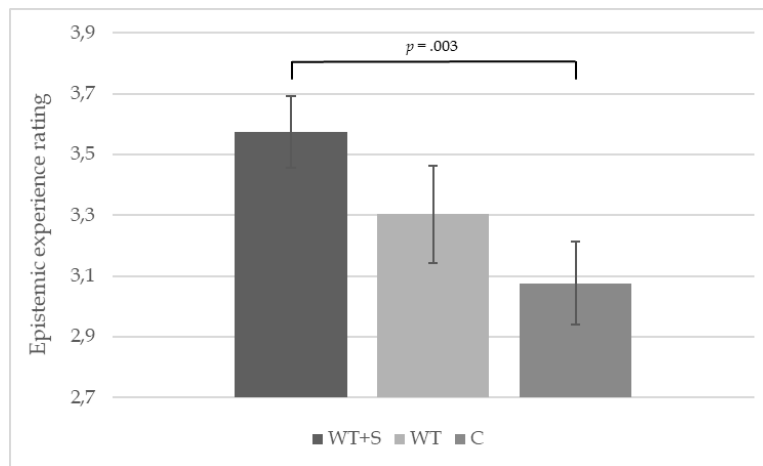


3.2 Epistemic Experience (RQ1-2b)

For epistemic experience, it was also Model 3 which fitted the data best, suggesting an interaction effect of Time and Condition (Figure 4.3). In the post-test, students in Condition WT+S rated their epistemic experience

significantly higher than the those in the Control Condition (mean difference: .492, $se = .149$, $p = .003$).

Figure 4.3. Epistemic Experience Ratings (Scale 1-5) in the Posttest, as Estimated Under Model 3



3.3 Effects on Course Content Knowledge (RQ1-2c)

For the two indicators of course-content knowledge, Model 1 fitted best, indicating an effect of Condition. Results are presented in Table 4.9. For number of new T-units, condition WT scored significantly lower than Condition WT+S (mean difference: -1.80, $se = .72$, $p = .013$) and than the Control Condition (mean difference: -2.11, $se = .77$, $p = .006$).

For mean length of new T-units, however, results were different. While the Control Condition scored significantly lower than Condition WT (mean difference: -2.62, $se = .97$, $p = .007$), there was no significant difference for Condition WT+S with the other conditions.

In summary, although students in Condition WT seemed to write less new T-units, their new T-units were relatively longer. While those in the Control Condition did write significantly more new T-units; those new T-units were relatively shorter.

Table 4.9. Estimated Means for Course Content Knowledge Indicators, Model 1

	Condition WT+S EM (SE)	Condition WT EM (SE)	Control Condition EM (SE)
Number of new T-units	6.41 (.43)	4.61 (.58)	6.72 (.50)
Mean length of new T-units	9.34 (.54)	10.35 (.73)	7.73 (.63)

3.4 Moderating Effects of Learner Variables (RQ3)

We tested whether learner variables moderated the intervention effect. For writing proficiency, there were no moderating effects of learner-variables. We did observe a better fit in Model 4, with the addition of cognitive engagement and self-efficacy (Appendix G), indicating a direct effect of those factors on writing-proficiency in all conditions. The effects were positive: students with higher self-efficacy scores, and students reporting more cognitive engagement beliefs, obtained higher holistic scores (Table 4.10).

Table 4.10. Estimates of Fixed Effects on Holistic Score for Two Covariates, Model 4

Moderators	<i>b</i>	<i>SE b</i>	95% confidence interval		<i>p</i>
			lower bound	upper bound	
Cognitive engagement	3.88	1.43	1.06	6.70	.007
Self-efficacy	.22	.11	.002	.44	.048

Furthermore, we have observed a moderating effect of emotional engagement and of self-efficacy, on *epistemic experience* (Appendix G, Model 7). In Condition WT and the Control Condition, these learner variables seemed to have no effect, while in Condition WT+S the two moderators did seem to contribute to the effect.

In condition WT+S, epistemic experience ratings developed differently for the different levels of emotional engagement beliefs: they remained unchanged from pre- to post-test for low-level emotional engagement beliefs, while they increased for high-level emotional engagement beliefs ($b = -.86$; $se = .32$, $p = .007$) (Figure 4.4). The same pattern was observed for self-efficacy (Figure 4.5). These results indicate, that the strategy instruction seemed more effective for students with high self-efficacy or high emotional beliefs levels, similarly to the aforementioned epistemic experience.

Figure 4.4. Epistemic Experience in Condition WT+S versus Control Condition for Students With Different Levels of Emotional Beliefs (= EB)

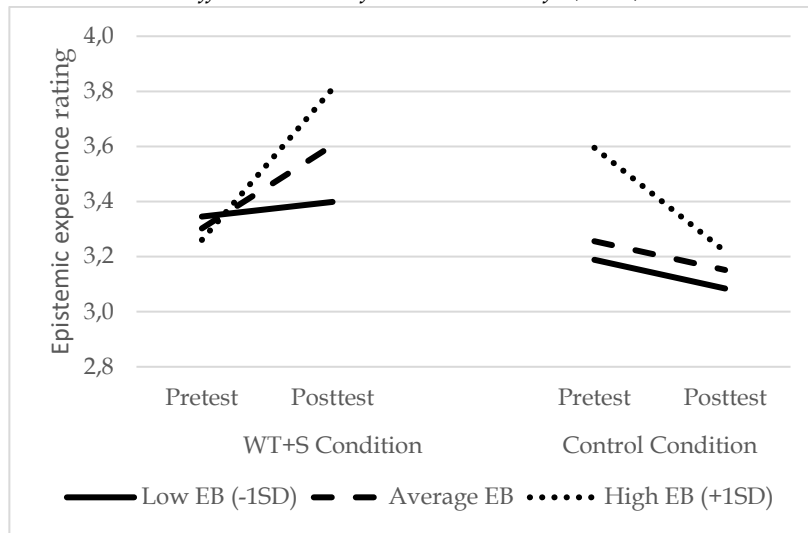
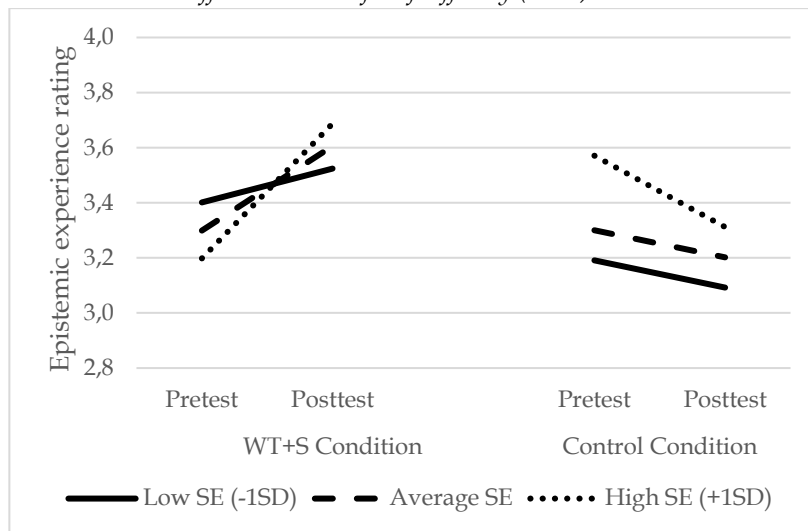


Figure 4.5. Epistemic Experience Condition WT+S versus Control Condition for Students With Different Levels of Self-Efficacy (= SE)



4. DISCUSSION AND CONCLUSIONS

In the current study, we aimed to investigate whether integrating writing tasks as learning activities into the history classroom, when accompanied by strategy instruction, would foster students' historical writing and history learning (pre-university education, grades 10-12). In order to do this, we proposed three research questions. Specifically, we have investigated whether integrating writing tasks as learning activities into the history classroom (RQ1), and additional discipline specific writing strategy instruction (RQ2), would improve the students' disciplinary writing proficiency, epistemic experience, and course content knowledge. Additionally, we have explored if certain learner variables moderated intervention effects (RQ3).

With regard to RQ1, we hypothesized that the replacement of workbook assignments by source-based writing assignments, would lead to greater epistemic experience in a transfer task, and more course-content knowledge in a recall task, but not to more advanced writing proficiency. Our results confirmed our hypotheses only in part. From the data, we can conclude that only replacing text book assignment by writing tasks (Condition WT), was indeed not sufficient for students to significantly outperform a non-writing control group on writing proficiency. These results confirm Monte-Sano's previous finding (2008), being that the act of writing alone was not sufficient for students to improve their evidence-based history essay writing. Furthermore, Condition WT did not outperform the Control Condition when assessing the epistemic experience in the transfer writing task, which was unlike our expectation. We had expected the students in the writing conditions, to experience the epistemic function of writing, especially after performing writing-to-learn tasks. Upon assessing course-content knowledge, students in the WT Condition wrote less new T-units in a post-test recall task, compared to the Control Condition, however, their T-units were significantly longer. Since longer T-units are associated with more clauses, implying more connections between elements, it could be conceivably concluded that the WT Condition was at least not inferior to the Control Condition with regard to course content knowledge. This finding seems in line with Klein's hypothesis (1999), that writing holds the intrinsic value of learning: writing in itself, is an instructive activity.

Concerning RQ2, we had predicted that those students who received additional writing-strategy instruction, would achieve higher text-quality scores than those students who performed writing tasks without process instruction, or than those students in a non-writing condition. We expected the strategy

instruction to reinforce the effects on epistemic experience, and course-content knowledge. From the data obtained, a possible conclusion could be that the inclusion of a discipline specific strategy instruction indeed effectuates an improvement of students' writing proficiency. Specifically, for Condition WT+S, students outperformed Condition WT and the Control Condition on *all* aspects of text quality: holistic score, text length, and discipline-specific criteria – situational understanding, multi-perspectivity, argumentation, source use, and structure. These results are similar to previous results of strategy instruction (e.g. De La Paz et al, 2017). Furthermore, Condition WT+S also scored the highest on epistemic experience.

With regard to course content knowledge, there were no differences observed between Condition WT+S and the Control Condition. The inclusion of strategy instruction and writing tasks, thus seems not to impact the course content knowledge, while, at the same time, it does seem a better option for obtaining disciplinary writing proficiency, and for a better epistemic experience, than merely implementing writing tasks.

When we include learner variables into the interpretation of the results (RQ3), two sets of results were observed: general and moderator effects. We found direct effects of cognitive engagement-beliefs and of self-efficacy, on holistic text quality; meaning, that students who believed that writing and thinking are intertwined, wrote better texts. A similar conclusion was drawn by other researchers (Villalón et al., 2015). A possible explanation for this, could be that students who consider writing tasks as learning tasks, perceive writing as more useful. However, we could not confirm this with the data; there were no significant correlations between cognitive engagement beliefs, and reported task effort. The general effect of self-efficacy on text quality, does confirm previous research outcomes (e.g. Sanders-Reio et al., 2014).

Furthermore, we found two moderator effects for epistemic experience in Condition WT+S only: emotional engagement and self-efficacy. Students with high emotional engagement beliefs and/or with high self-efficacy, seemed to benefit most from the strategy instruction; they showed more epistemic experience.

Although history teachers have generally tended to underrate writing process instruction, and might not consider this a priority, our results indicate that if teachers want to teach students to write historically, writing opportunities with writing process instruction is vital. The present study has shown that teachers can replace workbook questions on any topic, with evaluative writing tasks, and add strategy instruction. These additions not only increase

historical writing proficiency in transfer tasks, but also stimulate to experience the process of writing as epistemic, and expand students' course content knowledge. This is particularly important, as from a history teachers' perspective, a lagging growth of content-knowledge would be problematic. Providing process instruction is already often seen as a 'waste of time' for building knowledge, which would be even more so, if recall tests had confirmed this. However, our results have shown that writing tasks as learning activities were not harmful for course content knowledge at all. In fact, the recall test results showed that writing might even provoke more complex knowledge, as evidenced by the relatively extensive length of new units. Moreover, the results even suggest that students improved their historical reasoning, considering the higher scores on situational understanding, argumentation, and source use, which might be seen as indicators for this skill (Van Boxtel & Van Drie, 2018; Van Drie et al., 2014).

4.1 Strengths, Limitations, and Future Research

In this study, we aimed to overcome known obstacles for teaching literacy, such as the persistent 'pedagogy of telling'. The inclusion of writing-to-learn tasks was perceived as successful: the teachers of Conditions WT+S and WT experienced the writing tasks as instructive learning activities. During the evaluative interviews, the teachers indicated that they considered it 'the essence of history' to learn to think about complex historical issues from several perspectives. This makes writing evaluative texts a valuable activity for learning history, in a different way than the text book dictates.

Another obstacle to overcome, was teachers' lack of knowledge about writing processes. We did this, by using a solid instructional design with feasible tools (e.g., a video with modeling students), and by offering a PD-session in order to secure the correct and effective use of the intervention materials.

An important feature of the design of this study, was the flexibility of our intervention, in order to make the design feasible, and easy to implement, into the curriculum. We provided the teachers with many possibilities: teachers were not bound to a specific topic, a specific grade level, or to a narrow time-frame. Moreover, the students were not constrained by a fixed strategy either: we provided different routes, in order to accommodate each student's own writing routines. This research design also increased generalizability; the effects are not bound to a certain course within history education, as various thematic units were involved, in all conditions. Future research could focus

on the effectiveness of similar disciplinary approaches in other school subjects.

Although we consider adaptability an important strength of our instructions, it might also be seen as a limitation. The differences between groups, and thus also between conditions, may have resulted in different outcomes. However, we tried to maintain consistency between groups, and between conditions, during the intervention, by a) establishing and monitoring clear design principles for writing tasks and strategy instruction; b) using statistical analyses which level out group differences; and c) measurements, for example, the recall test contained a pre- and post-test; controlling for prior knowledge of the different topics. However, despite these precautions, groups remained unequal, for example with regard to the amount of time spent on content-knowledge building. Also, the groups differed with respect to duration of the intervention, implying different numbers of content lessons. Although this latter was perhaps to be expected, as topics differed from group to group, and additionally, topics were not equally difficult or complex.

Furthermore, the teachers' positive evaluations of the writing tasks, attest that they prefer discussing historical issues in depth over discussing many different topics superficially. In the writing conditions WT+S and WT, the topic has to be discussed in depth, since writing tasks provoke this. However, the approach of covering a large set of topics on a very shallow level ("a mile broad, an inch deep"), is quite common in history education, and this thus might have been the dominant approach in the control condition. This could be seen as a limitation, since there is no clear view on the interference between complexity of the topic and time spent. Moreover, we could question the feasibility of meeting all curriculum demands with this in-depth approach, in the long term. This concern could, in turn, restrain teachers from applying writing tasks more widely.

Another possible limitation is that the recall test was used as a surrogate for course content knowledge. Previous research has criticized recall tests as a measure of knowledge gain through writing, since this test is likely to detect knowledge accretion, which is only a peripheral outcome of writing (Schumacher & Nash, 1991). However, the recall measure was not the only content measure in our study; in the assessment of the students' texts, we deliberately included content-related indicators, such as argumentation, as well.

A final limitation worth mentioning, is that the data from the current study, is inconclusive about the potential durability of the effect observed. In

short, it has not been made clear if a writing task later on in the school year, would need to be accompanied by another process instruction, or if a quick reminder of the strategy would be sufficient. This question of durability of the process regarding instruction for students' disciplinary literacy, should be further examined in future research.

4.2 Conclusions

In conclusion, after only two strategy lessons and two writing opportunities for practice, of about 50 minutes each, students overall wrote better texts. Compared to the WT Condition, where students had similar writing opportunities, but were uninstructed, the added process instruction made a significant difference. Moreover, in the setting in which there were writing opportunities, but without the instructions, outcomes were comparable to those for students in a control condition without writing. Although history teachers generally tend to underrate writing process instruction, and do not consider it a priority, our results underline the necessity, and the importance, of including writing strategy instruction into their discipline anyway; and contradict the archaic paradigm, that including writing tasks and process instruction are a 'waste of time'.

CHAPTER 5

FOSTERING PHILOSOPHY TEACHERS' DISCIPLINARY WRITING PRACTICE: A MULTIPLE-CASE DESIGN STUDY*

In this design study, an instructional unit open to contextual modifications was designed with the aim of fostering secondary school students' philosophical writing. Since teachers' teaching practices have been found to be guided by their beliefs and conceptions of teaching, the implementation of the design was prepared and supported by guiding sessions that stimulated philosophy teachers' thinking about writing and writing support in their discipline.

Based on design principles for effective writing tasks, three philosophy teachers designed innovative source-based writing tasks to use in their 10th grade classes. They subsequently implemented the tasks and provided additional discipline-specific writing strategy instruction, which was designed by the research team based on previous writing research.

Our study focused on teachers' interactions with the instructional design with the goal of testing the resilience of the design in open, ecologically valid situations. Furthermore, we aimed to investigate whether the unit was effective for students' development of philosophical writing.

A multiple-case approach using predominantly qualitative measures suited our research aims. After teachers had implemented the instructional unit in their own context, we conducted evaluative interviews with teachers and students concerning contextual adaptations. An external jury analyzed students' texts to determine their actual learning achievements. Teachers' insights into student progress were obtained from reflective interviews that featured comparisons between the observed and expected results.

* Chapter 5 is based on: Holdinga, C. C., Van Drie, J. P., & Rijlaarsdam, G. C. W. (2023). Fostering philosophy teachers' disciplinary writing practice: A multiple-case design study. Under review.

The results showed that teachers integrated the design with their specific contexts within the design parameters and judged the design to be feasible, valid, and effective for students' philosophical writing development. The experience of writing task design and the implementation of writing strategy instruction drove teachers to question the roles that writing tasks might play in philosophy education, the actual requirements of high-quality philosophical writing, and the possible value of writing strategy instruction in their discipline. These contemplations indicate changes in belief, which are necessary for genuine improvement in teacher practice. We therefore conclude that the findings indicate encouraging results regarding the effectiveness of an instructional unit accompanied by teacher guidance sessions that can promote reflection on the development of students' writing.

1. INTRODUCTION

The reading and writing skills that are necessary for the twenty-first century have increased. Today, literacy requires the ability to use reading and writing to acquire knowledge, solve problems and make decisions academically, personally and professionally (Goldman, 2012). However, according to recent international reports, Dutch secondary school students' literacy skills have exhibited an alarming trend of decline (OECD, 2018). Similar trends have been reported in other countries. In the United States, for example, recent national assessment reports have shown that the reading performance of US high school students has not improved since 1971, with only 38% of high school students scoring at or above the level of proficiency (Goldman, 2012). The impoverishment of students' literacy skills is problematic: these skills are a prerequisite for independent participation in our highly literate society. Moreover, literacy skills are crucial for students' subject matter learning and cognitive development (Graham et al., 2020).

The common view of the development of students' literacy in recent educational research has been that it is the responsibility of every teacher, not merely that of teachers in the language department or L1 teachers. Researchers have also noted that developing literacy is not merely a matter of developing proficiency in general reading and writing practices regardless of content (Moje, 2008; Shanahan & Shanahan, 2008). Instead, literacy is now viewed as a crucial element of the task of supporting students in understanding and developing knowledge of various subjects. Therefore, literacy instruction

should be treated "as a key part of the broader effort" (Heller & Greenleaf, 2007, p. 1).

In the subject of philosophy, which is the context of this study, students are asked to perform a variety of linguistic tasks, which demand a proficient level of reading and writing skills. Students are required to read and interpret complex primary philosophical texts with the aim of enabling them to answer philosophical questions. Reading philosophical texts is quite difficult: these texts contain unfamiliar vocabulary, abstract ideas, complexly organized writing, and unsettling views (Concepción, 2004). In their search for a satisfactory answer to a philosophical question, students often read multiple texts to compare different philosophical views. Ultimately, students adopt a reasoned position toward the issue at hand. To demonstrate their proficiency in philosophical reasoning, students can be asked to express the position they support in, for example, the form of a Socratic conversation or a written text.

These activities also rely heavily on students' literacy skills. It might therefore seem obvious for a philosophy teacher to address students' literacy skill development; however, such a focus is not always the case in educational practice. In a previous study on teaching philosophical literacy (Koek, 2020), 90% of philosophy teachers responded that developing their students' literacy was a challenge. Furthermore, only 25% of the respondents reported that they explicitly explained to their students how to read a primary philosophical text. We thus conclude that there are opportunities to improve educational practice regarding disciplinary literacy development. With this study, we aimed to take advantage of this opportunity by developing, implementing, and evaluating an instructional unit intended to enhance students' philosophical reading and writing.

1.1 Teaching Disciplinary Literacy

The integration of literacy into classroom content places a heavy burden on teachers. Teachers who are accustomed to teaching a subject are required not only to have knowledge of the relevant content (i.e., facts, concepts, structural relations, reasoning, and argumentation) but also to obtain knowledge of the reading and writing procedures specific to their discipline (Monte-Sano et al., 2014). Moreover, research conducted by Goldman and colleagues (2016) produced a framework featuring five core concepts on which disciplines differ: (1) epistemology; (2) ways of enquiry and reasoning; (3) concepts, principles, and frameworks for describing and analyzing phenomena; (4) types of sources; and (5) text genres and language use. These concepts are clusters of

the types of knowledge that enable subject teachers to formulate learning goals that are targeted at what students need to know and to be able to do to attain high levels of literacy in their disciplines. This task might be quite a challenge for teachers who have not been educated to provide instruction on these aspects of literacy.

Teaching materials aimed at the development of students' philosophical reading and writing with the goal of supporting teachers' practice are scarce. For other subjects, enhancing disciplinary literacy has been researched more extensively. For example, this topic has been explored with regard to the subject of history by De La Paz and colleagues (e.g., De La Paz & Felton, 2010; De La Paz et al., 2017; Monte-Sano & De La Paz, 2012) and in the fields of math and science by Hand's research team (e.g., Chen et al., 2016; McDermott & Hand, 2016; Villanueva & Hand, 2011). In response to this lack of research, we developed an instructional unit, which was inspired by De La Paz's research on the development of historical writing and is tailored to the subject of philosophy, with the aim of enhancing students' philosophical literacy. A distinctive aspect of our study was our focus on discipline-specific aspects of philosophical writing processes and texts.

1.2 Enhancing Classroom Practice

To promote proficient student writing, qualified writing tasks and instruction are necessary. This situation requires teachers to develop knowledge about writing processes and the teaching of writing as well as to reflect on their beliefs concerning the role of writing and the characteristics of disciplinary texts. Teachers who experience effective professional development obtain new knowledge, which changes their beliefs and classroom practice, thus ultimately fostering student learning (Desimone, 2009). The professional development of teachers is thus key to improving student achievement.

Previous studies (Samuelowicz & Bain, 1992) have distinguished among five qualitatively different conceptions of teaching, from learning-oriented on the high-quality side of the scale to content-oriented on the low-quality side. Teachers' conceptions of writing are highly correlated with their teaching approaches (Kember, 1997). Teachers who view teaching as learning-oriented are likely to include more project-based learning activities, in which context students are responsible for their own learning process. In this case, the role of the teacher is to encourage and monitor students' work. However, teachers who view teaching as content-oriented are more likely to focus on imparting knowledge. Ho and colleagues (2001) have recommended a developmental

approach that focuses on conceptual change; these authors argued that real improvement in teachers must begin with a change in the way they think about teaching or, in this case, about teaching writing. Therefore, we designed a teacher guidance program to stimulate contemplations of writing tasks, instruction, and support.

According to the philosopher of science De Vries (1984), research can be practically relevant in a technical way by generating new techniques and ways of doing; however, it can also be relevant in a cultural way by generating new ways of expressing, viewing and thinking. Both types of relevance are addressed in this study, as teachers changed their practice and challenged their beliefs and attitudes. We consider these cultural changes to be important with regard to their long-term impact on teacher practice.

1.3 Research Questions

The primary goal of this study was to equip teachers to integrate qualified disciplinary writing instruction into their practice. We conducted a design study that consisted of three phases: (a) analysis and exploration; (b) design and construction; and (c) evaluation and reflection (McKenney & Reeves, 2019). In phases a-b, we focused on preliminary research questions aimed at both designing an instructional unit that can enhance students' philosophical literacy and providing accompanying teacher guidance to ensure the success of the integration and adoption of the unit:

RQ1: Which design principles can be derived from previous research on the development of secondary school students' philosophical literacy?

RQ2: How can these design principles be developed into an instructional unit?

In the third phase, the instructional design was implemented, evaluated, and reflected on. In guiding sessions, three teachers were prepared for writing task design and the integration of this form of instruction into their 10th grade classes.

We explored teachers' interaction with the materials with the goal of understanding how they employed and examined learning activities drawn from the innovative instructional design as they worked to improve their students' disciplinary writing in response to philosophical issues. The purpose of this exploration was to determine the extent to which teachers' instructional

practices had actually changed and the ways in which teachers and students perceived these changes. The third research question was therefore as follows:

RQ3: How do philosophy teachers interact with (a) principle-based writing task design and (b) an instructional unit intended to support students' disciplinary writing?

Next, we aimed to obtain insights into the perceived effectiveness of this approach with regard to students' writing proficiency and teachers' conceptions of student progress in philosophical writing. RQ4 was therefore as follows:

RQ4: To what extent does the intervention contribute to (philosophy teachers' conceptions of) progress in students' philosophical writing development?

With regard to RQ4, we asked teachers to define the expected level of student performance, i.e., the performance level that they thought their students were likely to achieve according to their expectations, and we then confronted teachers with the actual level of students (Rijlaarsdam & Janssen, 1996) based on independent assessments of students' writing.

Given the nature of RQ3-4, in the attempt to obtain a profound understanding of teachers' practices and beliefs, we considered multiple-case research featuring predominantly qualitative measures to be an appropriate research approach.

2. ANALYSIS AND EXPLORATION PHASE

2.1 Exploration of the Context

We explored the educational context in a prestudy to ensure that the intervention was suitable for (a) the regular philosophy curriculum, (b) teachers' views on the level of philosophical writing required from students, and (c) teachers' needs for guidance with regard to providing writing instruction. Therefore, we conducted interviews with 11 philosophy teachers drawn from different schools. Interviews took approximately one hour each, were audiotaped and were subsequently transcribed. We asked teachers to bring an example of a writing task that they had recently used at the upper-secondary level as well as two exemplars, including one weak exemplar and one strong exemplar. Five topics were discussed in the interview: writing task character, assessment

criteria, support practices, cognitive processes, and writing beliefs. For the full instructions for the interviewees and the interview guide, see Appendix E. The results showed that teachers used tasks that were mostly aimed at stimulating philosophical thinking. For example, students were asked to write a philosophical essay or an elaboration of a thought experiment. Teachers were accustomed to providing feedback on texts and regularly asked students to write several drafts. They varied in terms of their satisfaction with their students' writing level; however, all teachers indicated that not all students reached the required level. They reported finding it difficult to guide these students in this regard. The provision of process instruction or process feedback was scarce. Teachers seemed to lack knowledge about writing processes, as they struggled to make explicit the cognitive activities that they assumed students would perform. In conclusion, the starting situation was that while teachers were accustomed to employing writing assignments in their philosophy teaching, they lacked the knowledge and tools necessary to instruct students or support them in the task of writing.

2.1 Design Principles

In Chapter 3, we established design principles for literacy development in history. Based on the prestudy and a literature search, we adapted these design principles for the instructional unit to the discipline of philosophy:

If we want students to develop a profound understanding of philosophy through writing, then it is best

- 1. to use writing tasks that can prompt a discussion of a concise philosophical issue and to accompany this prompt with various (primary) sources that represent multiple perspectives on the issue at hand; and*
- 2. to provide students with discipline specific, dual-route, reading-writing strategy instruction that is easily applicable for teachers.*

We opted to focus on writing-to-learn tasks since content focus is considered to be an effective feature in educational innovation (Van Veen et al., 2012). In the Netherlands, the main aim of philosophy education is to teach students how to philosophize in their own right. Essay writing can contribute to this goal; it encourages a writer to "explore" and progress through an entire "train of thought" (Velema & Groza, 2020). Philosophical essays are common tasks

that teachers use as a learning activity (Marsman, 2010), thus enhancing the feasibility of our design. Therefore, it seems to be best to design tasks that feature a philosophical issue as a prompt: an open question that provokes a process of thinking and that cannot be solved by empirical research or observation. Since reading and writing may reinforce each other (Graham et al., 2018), we recommend the use of tasks based on sources. According to previous research on source-based writing, primary sources containing multiple views on the issue evoke more 'sourcing' and 'referencing' (Britt & Rouet, 2012).

Our prestudy showed that philosophical essay tasks in upper secondary grades generally result in long texts: a length of 800-1000 words is not uncommon. Nevertheless, for two reasons, we aimed to emphasize concise tasks, resulting in short texts. First, we aimed to ensure that students wrote within sight of the teacher to facilitate the provision of process support. With short tasks, this goal is feasible. Second, we aimed to enable students to master a procedure and to provide them with multiple opportunities to practice this procedure. Shorter tasks would be most efficient in this case; writing multiple long texts would be tiring and demotivating for students.

We derived the second principle from research on general writing instruction. Providing direct explicit instruction and supporting students' writing processes are known to be effective methods of writing development in general (Graham & Harris, 2017). In another study that focused on the field of history (Chapter 4), we concluded that strategy instruction is also effective for disciplinary writing. In addition, research (Kieft et al., 2007) has shown that adapting writing instruction to students' writing strategies is effective for writing-to-learn, which led us to employ a dual-route strategy.

Previous research has shown that innovative methods should be easily applicable for teachers. Demands regarding practicality (instrumentality, congruence, and low cost; Westbroek et al., 2020) should thus be taken into account. Accordingly, the unit should be open to contextual modifications. When the design is flexible with regard to adaptation to different topics and levels of performance, it drastically enhances the feasibility of this approach for teachers.

Finally, our prestudy showed that philosophy teachers did not have much experience in teaching writing. Therefore, we require guiding activities to support the integration and implementation of the unit into teachers' practice. To guide the design, we considered critical features of effective professional development (Desimone, 2009; Van Veen, et al., 2012): content focus, active

learning, alignment with teachers' goals, and sufficient duration. A final design principle thus focused on teacher guidance:

If we want teachers to design qualified writing tasks and provide writing strategy instruction, then it is best

3. *to support teachers' integration and implementation process with guiding activities that can prompt reflection on the part of teachers while taking into account critical features of effective professional development.*

3. DESIGN AND CONSTRUCTION PHASE

3.1 Writing Task Design

For the benefit of ecological validity and teacher engagement, we entrusted task design to the participating teachers. Therefore, the first principle was operationalized into three requirements: the writing tasks were required to (1) prompt an exploration of a concise philosophical issue, (2) be based on various (primary) sources and represent multiple perspectives, and (3) be appropriate for the duration of one class.

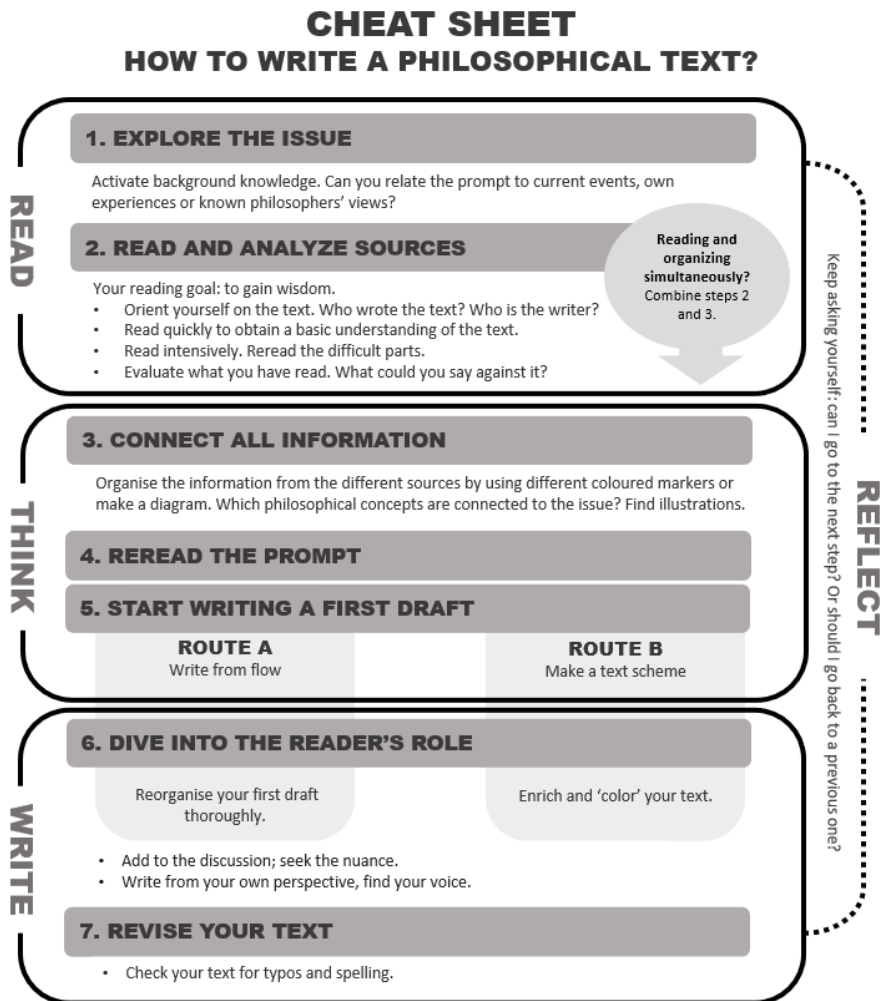
3.2 Writing Strategy Instruction

To elaborate the second principle, we designed (a) a strategy for students that explains how to perform the writing tasks most effectively and (b) an instructional design aimed at teaching students the strategy effectively.

3.2.1 The RTW Strategy

Based on a literature search, context exploration, and previous studies (Chapters 2, 3 and 4), we designed a genre- and discipline-specific Read-Think-Write strategy (RTW strategy). The RTW strategy is divided roughly into three main cognitive processes: reading, thinking, and writing. Furthermore, students are prompted to monitor the entire process since the three processes overlap and interact to a large extent. We aimed to address all five core concepts in which disciplines differ according to the framework developed by Goldman et al. (2016). For example, students were prompted to connect abstract concepts with concrete examples (addressing overarching philosophical frameworks) and to find their own voice in their texts (addressing philosophical epistemology/discourse language). Figure 5.1 shows the strategy as presented to students.

Figure 5.1. The Strategy as Presented to Students (Concise Translation; the Original "Cheat Sheet" Was More Extended and Was Presented on Folded A5-Wrapper in Full Color)



The first step in the seven-step strategy is to explore the issue presented and to activate prior knowledge (Corcelles Seuba & Castelló, 2015). Subsequently, students start reading the source materials to obtain enlightenment: to understand and reflect on what has been stated fully (Concepción, 2004) (Step 2). The phase of reading is aimed at obtaining a situation model: an elaborated

interpretation of the issue presented as described by the source (Britt & Rouet, 2012; Rouet & Britt, 2011). Creating such a model requires comprehension of content, reprocessing, recursive actions and a high standard of coherence (Van den Broek & Helder, 2007).

The next step is to connect the sources (Step 3) (Borren, 2012). This stage could partly overlap with the second step, as it might be possible for some students to connect information immediately with background knowledge or information they have previously read while reading. This step, which involves organizing source information, should lead to an integrated mental model: an internal representation of the issue discussed across texts (Britt & Rouet, 2012).

Step 4 is a reminder of monitoring. Subsequently, at the latest, students start writing at Step 5. The goal of writing during this phase is to stimulate knowledge constitution (Galbraith, 2009; Galbraith & Baaijen, 2018). The knowledge constitution process activates and articulates knowledge that is implicit or still diffuse. We offer two routes for writing that appeal to different writing preferences.

A first route is to write freely (Route A), i.e., using an intuitive style of writing. The free writer in fact follows Elbow's (1973) dual drafting strategy. In the first phase of writing (Route A, Step 5), the focus is on generating a first draft of the text without paying excessive attention to the rhetorical aspects that the text should or should not contain. In the second draft phase (Route A, Step 6), the writer focuses on the text itself, the language, the rhetoric, and the structure. Since we focus on short tasks, the drafts will not be actual drafts in the sense intended by Elbow. Instead, drafts progress in rapid succession.

A second route is preplanning (Route B), which involves a more decisive style of writing. Planning of the text (Route B, Step 5) might be performed at a local level - planning sentence by sentence - or at a global level - writing based on a global text plan. Creating a mind map or organizing information with the help of a text scheme is an alternative means of generating ideas. In the subsequent phase (Route B, Step 6), the text is expanded: students are instructed to "put flesh on the bones".

In general, Step 6 involves exploring the reader's perspective and considering the text critically. We then instruct students to organize the knowledge in a process of knowledge transformation (Galbraith & Baaijen, 2018). This process is necessary to transform students' knowledge into a rhetorically effective form. Finally, Step 7 focuses on editing.

3.2.2 Instructional Design

To teach students the strategy, we followed the instructional model proposed De La Paz and Felton (2010) to develop students' writing proficiency in the field of history. This model was inspired by the 'classic' model of Self-Regulated Strategy Development (SRSD) presented by Harris and Graham (1996). The framework for instruction contains five stages: develop background knowledge, describe it, model it, support it, and independent performance. We adopted these stages and provided them with content in terms of learning activities. In the current study, the intervention consisted of six lessons, which are described in Table 5.1.

A teaching manual was developed to convey the instructional unit to the teachers. For the intervention lessons, students received a paper workbook containing a first writing task (T1), Lesson 1, and Lesson 2. The writing tasks were to be completed using a computer.

The instructional unit starts with a writing task (T1), and the students' goal is to "experience" the task. This experience functions as a "hook" for Lesson 1, when students reflect on their performance.

Subsequently, students are presented the Read-Think-Write strategy. The strategy is explained by the teacher and modeled by peers in a video as a mode of observational learning (Braaksma et al., 2004). During this video, explanatory parts are alternated with four modeling peers, who were 12th grade students from other schools (Figure 5.2).

Figure 5.2. Stills from the Video with Modeling Peers (Translated; the Original Video Was in Dutch)

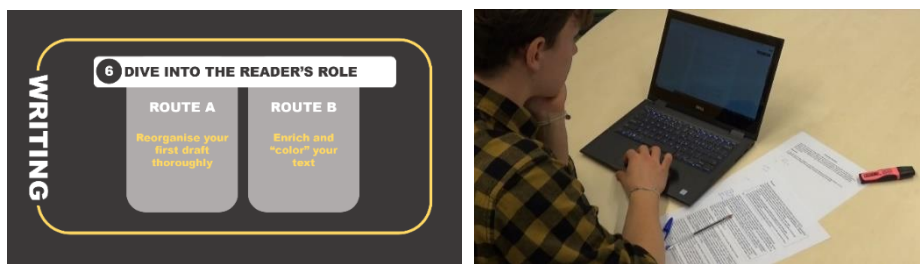


Table 5.1. Description of the Design-as-Constructed

Session	Stage	Learning activities	Description
Writing task (T1)	<i>Develop background knowledge</i>	Task experience	Students "experience" the whole task; they perform a first evaluative task.
Lesson 1	<i>Describe it</i>	Reflecting on the experience	The students write down what they thought was easy while performing the task and what was difficult for them.
		Building new knowledge Comparing the strategy to one's own experience	The teacher presents and explains the Read-Think-Write strategy. Students compare their task experience to the strategy presented.
		<i>Model it (process)</i> Observing the strategy demonstrated by a modeling peer (video) Relating to the performance of others	Students watch a video together on the main screen. This video shows modeling peers, who demonstrate how each step of the strategy can be performed. As a processing activity, the students individually note the elements of the strategy that are most useful to them.
Lesson 2	<i>Model it (product)</i>	Analyzing/assessing peers' texts	The teacher selects two-three texts written in response to T1 to use as exemplars. Students individually note positive and negative aspects of each exemplar.
		Generating criteria	In a class discussion of the exemplars, students generate a list of criteria.
		Applying new learning	Students apply the criteria to a text fragment of their choice by rewriting/revising that fragment.
Writing task (T2)	<i>Support it</i>	Scaffolded practice	Students perform T2, practicing the strategy. This practice is scaffolded with the support of the teacher and the written guide (strategy cheat sheet).
Lesson 3	<i>Support it</i>	Providing peer feedback	Students bring the text they wrote in response to T2 to class and exchange their texts in a group of four students. They provide each other with feedback.
Writing task (T3)	<i>Independent practice</i>	Individual practice	Individual practice with the help of the written guide.

These peers were recruited to perform T1 on camera and filmed anonymously from an over-the-shoulder perspective. There was no script; students performed the task naturally in their preferred manner in approximately 40 minutes. One researcher (LH) selected scenes that were suitable for illustrating how to perform each step of the strategy and each "route", ultimately creating a video of 12 m 46 s in length. The illustrations contain scenes in which students struggle and engage in monitoring. For example, they conclude that they did not understand a source and start rereading an excerpt.

Students performed a compare-contrast assignment while watching the video to help them reflect on their own process. In the teaching manual, teachers are instructed to pause the video after each stage (read-think-write) and to reflect on the models in a class discussion.

The design-as-constructed includes two feedback lessons. In Lesson 2, teachers are instructed to select two-three student texts written in response to T1 to discuss in class, and in Lesson 3, students discuss the texts they wrote in response to T2 with their peers in small groups consisting of three-four students. Analyzing and discussing exemplars is a means of conveying teachers' tacit knowledge about what criteria actually mean (Polanyi, 1973). In previous research (Orsmond et al., 2002), the discussion of exemplars has been proven to reduce differences between students' and teachers' assessments.

While completing T2, students can collaborate with their peers (Corcelles-Seuba & Castelló, 2015) and are supported by their teacher. At T3, students write independently, supported only by a written guide (the "strategy cheat sheet") (Martínez et al., 2015).

3.3 Teacher Guidance Activities

We designed a teacher guidance program that fit the design requirements following the prestudy and the recommendations for professional development for teachers (Desimone, 2009; Van Veen, et al., 2012). An overview of activities is presented in Figure 5.3.

The program was designed to feature two objectives. The first such objective is to support the integration and implementation of the instructional unit into teachers' practice. Initially, Session 1 was a one-on-one briefing (teacher and researcher) that enabled teachers to take note of design principles for effective writing-to-learn tasks and to understand the structure of the instructional unit and the underlying rationale. Based on this session, teachers could become well prepared for the implementation process in an authentic setting:

their own classroom. This session thus met the recommendation of content focus.

Subsequently, teachers designed their own writing tasks to use in class. The research team provided feedback when requested. Teacher involvement in the design process enhanced active learning and coherence with teachers' goals. Subsequently, teachers implemented the instructional unit and reflected on their implementation process in an evaluative interview (Session 2). Since the activities were spread over one semester, the duration requirement was met. Although collaborative participation is a fifth critical feature of effective professional development, we considered an individual trajectory to be appropriate at this stage focusing on the beginning of innovation.

Our second goal was to equip teachers to implement high-quality writing instruction and support in their future teaching. Therefore, we aimed to stimulate teachers' reflection. We included activities aimed at prompting contemplation of writing instruction and writing quality. We asked teachers to select for each task a benchmark text that they considered to be average in their group. This activity aimed to uncover teachers' expected level of progress at the group and student levels. In Session 3, teachers were asked to reflect on students' actual level of performance as determined by jury teams who assessed students' texts.

4. EVALUATION AND REFLECTION PHASE

4.1 Research Design

Three philosophy teachers implemented the instructional unit in their 10th grade classes to develop students' philosophical writing. We investigated teachers' interaction with the design elements (tasks and instruction) by monitoring and evaluating the implementation process with the goal of answering RQ3 and RQ4.

To explore the effectiveness of the intervention for student learning, we used a semiexperimental design on three measurement occasions (T1-T2-T3); teacher-designed writing tasks functioned as measurement occasions at T2 and T3. This approach thus resulted in different measurements for each participating teacher at T2 and T3. T1 was similar for all groups and was developed by the research team.

Independent jury teams assessed students' texts and teachers' tasks. Subsequently, teachers were asked to contemplate students' text scores. The

results of the study are analyzed and presented as a multiple case study (Yin, 1994) using the teachers as cases.

4.2 Participants

Three philosophy teachers (Teachers A, B and C) from three different schools participated in this study, each focusing on their own preuniversity 10th grade class (Groups A, B, and C, respectively). All teachers also participated in the prestudy. The teachers participated voluntarily and were interested in the implementation of an intervention aimed at the development of students' disciplinary writing. The teachers had between six (Teacher B and C) and eight (Teacher A) years of experience teaching philosophy. All three teachers had obtained a master's degree in philosophy and had a teaching qualification (master's degree) for teaching philosophy in secondary education.

In Dutch secondary schools, philosophy is not a mandatory subject. At the schools of Teachers B and C, philosophy was taught beginning in 10th grade as a subject of choice. In Teacher A's school, philosophy was also taught in 7th and 8th grade as a mandatory subject for all students. Beginning in 10th grade, it was optional. All three groups thus consisted of students who chose philosophy as a subject; however, only students at School A might be expected to have a more extensive knowledge base.

In total, 56 students (age: 16-17) participated (group sizes: A 22; B 17; C 17). Students actively consented to participate in this study; one student from Group C objected. Students' parents were informed of the study via regular communication channels; no parent objected.

4.3 Procedures and Data Sources

Data sources were interwoven with procedures to guide teachers' implementation. An overview of implementation activities, research activities and data sources is provided in Figure 5.3.

Figure 5.3. Overview of Teacher Guidance Activities, Research Activities and Data Sources

Teacher guidance activities	Research activities	Data sources	RQ
	Prestudy: Exploration of context Designing the instructional unit	<ul style="list-style-type: none"> ➤ Explorative interviews 	RQ1-2
SESSION 1 Briefing Writing task design Implementation of instructional unit SESSION 2 Evaluative interview	Preparing teachers for implementation Providing feedback Check for implementation fidelity Implementation assessment with a. teachers and b. students	<ul style="list-style-type: none"> ➤ Teacher-designed writing tasks ➤ Classroom observations ➤ Teacher logs ➤ Evaluative interviews with teachers ➤ Evaluative interviews with students 	RQ3
Defining the expected level External assessment of tasks and actual level of student achievement SESSION 3 Reflective interview	Teachers select benchmarks Jury teams assess tasks and texts Evaluation of student progress	<ul style="list-style-type: none"> ➤ Selected benchmark texts ➤ Teachers' explanations ➤ Field notes of jury team sessions ➤ Text scores ➤ Reflective interviews with teachers 	RQ4

4.3.1 Session 1: Briefing

With each teacher, we organized a one-on-one session (lasting approximately one hour) to explain all aspects of the teaching manual. We highlighted the essential elements of the intervention and explained the rationales underlying the design in general. We hypothesized that this briefing would improve the quality of the implementation. Subsequently, each teacher was instructed to design two writing tasks tailored to their regular year plan based on our design principles.

4.3.2 Writing Tasks

As measurements for students' writing proficiency, we used three writing tasks (T1, T2, and T3) that were fully integrated into the instructional unit to avoid spending lesson time solely on research purposes.

T1 was developed by the research team; this task was the same for all groups and functioned as the foundation of the strategy instruction. Furthermore, it served as a baseline for assessing change over time. T1 did not rely excessively on philosophical concepts or an extensive philosophical knowledge base; it was intended for use in all participating groups and did not require prior philosophical knowledge.

T1 asked the following question: "To what extent are humans capable of true altruism?" Four source texts were provided (mean length: 117 words) to represent different perspectives on the issue. The first source text (123 words) was an excerpt from a blog written by social psychology professor Roos Vonk, who indicated that people may have all kinds of egoistic motives to help others. In the second text (106 words), journalist Matt Ridley discussed Richard Dawkins' theory of selfish genes: we would never have children if we did not have genes that were selfish. In the third text (113 words), philosopher Adam Smith claimed that even "the biggest villain" cares about other people's feelings. In the fourth source text (127 words), Buddhist writer and monk Matthieu Ricard cited a study proving that children act altruistically at a young age, thus suggesting that altruism could be "innate". The writing tasks for measurement occasions T2 and T3 differed from group to group since these tasks were designed by the participating teachers.

Students wrote their texts for T1, T2, and T3 on a computer in a computer room (Group A), using iPads (Group B), or on laptops (Group C).

4.3.3 Session 2: Evaluative Interviews

Immediately after the intervention, we conducted evaluative interviews with teachers and students to obtain insights into their interaction with the design and implementation quality. Teacher interviews took approximately one hour. Interviews with students were conducted in small groups to obtain information regarding the fidelity and validity of the intervention materials as well as students' self-perceived progress in writing. Five students from school A (two female) were interviewed, as were three students from school B (all female) and three students from school C (one female). Student interviews took approximately 45 minutes. All interviews were audio-recorded and subsequently transcribed into written protocols. Interview guides are presented in Appendix E.

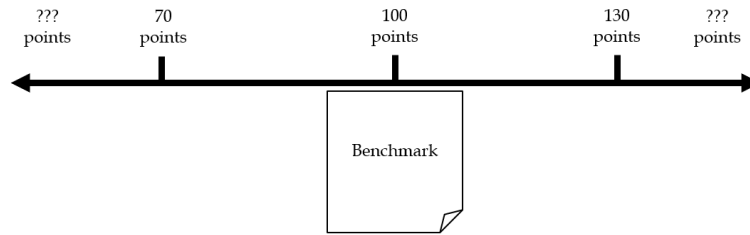
4.3.4 Benchmark Selection

After the implementation of the instructional unit, we asked each teacher to select for each assignment (T1, T2, and T3) one text from the written texts in their class that they would label as a benchmark text: an "average" text representing the average philosophical writing proficiency of students in the group. We invited teachers to elaborate on their decisions: we asked for explanations of the strengths and weaknesses of each benchmark text and suggestions for improvement of the text. Teachers performed this task individually and communicated their findings via e-mail. This activity was aimed at encouraging teachers to contemplate the criteria they used for text quality. From a research perspective, this activity provided us with insights into teachers' expected level of student achievement.

4.3.5 Jury Team Sessions

Independent jury teams analyzed the texts written by students at T1, T2 and T3 to determine students' actual level of writing proficiency. The selected benchmark texts functioned as points of reference at a score of 100 points (Figure 5.4). Per assignment, all other texts were assigned points in comparison to this benchmark. For example, a text could score 80 points if it was judged to be of lesser quality or 120 points if its quality was rated higher.

Figure 5.4. Assessment Scale



Jury teams consisted of one (Group A) or two (Group B, C) philosophy teachers in addition to a researcher (LH) who participated in all three of the teams. Each jury team collaboratively read and analyzed all three tasks of one group, discussed the benchmark texts and the accompanying explanations of the intervention teacher, and determined a score for three to four texts produced by students on every topic. The aim of the discussion was to obtain a “shared understanding” of the relevant criteria and standards. T1, T2 and T3 were discussed in random order. The collaborative sessions took approximately 90 to 120 minutes each. The researcher took field notes during the sessions.

The remaining texts were individually assessed by the jury team members. Reliability coefficients for the individually assessed texts were .89 for the jury team for Group A, .91 for Group B, and .92 for Group C. Further analyses were conducted by reference to mean scores.

4.3.6 Session 3: Reflective Interviews

After the text rating procedure, we conducted reflective interviews with Teachers A, B, and C. Each interview consisted of five parts, most of which were supported by visual representations of the group results. We started with a review of the designed writing tasks and the teacher's expectations regarding their students' results on those tasks (Part 1), taking into account the level of the selected benchmark texts. Subsequently, the results for each task were presented in three graphs, one for each task; each graph contained all students' individual results for that particular task (Part 2). Then, we presented the group results in the form of an overall line graph showing group progress (Part 3) as well as in a quadrant visualizing individual students' progress from T1 to T3 (Part 4). Finally, we discussed the students whose results were atypical (Part 5): we showed the results at T1, T2, and T3 for a selection of students whose scores either highly increased (+1 SD or +2 SD) or highly

decreased (-1 SD or -2 SD) from T1 to T3. For each graph, we asked teachers (a) whether the results were consistent with what they had expected at the group and/or individual level; (b) to explore possible explanations for the results; and (c) to respond to trends and abnormalities at the student level. The reflective interviews took approximately one hour, were audiotaped and were subsequently transcribed into written protocols.

Based on the analyses of the interviews, themes were extrapolated. For RQ3, we distinguished among four themes: teachers' interactions with (1) writing task design, (2) the strategy, (3) the modeling video, and (4) the discussion of exemplars. For RQ4, five themes were differentiated: (1) the characteristics of philosophical writing, (2) the evaluation (difficulty/length) of designed writing tasks, (3) benchmark quality, (4) expected progress, and (5) perceptiveness. Per theme, all relevant data sources were involved in the elaboration (Figure 5.3).

5. RESULTS

In the following sections, we present a cross-case analysis based on evaluative interviews, jury team discussions and reflective interviews with the goal of providing a more general perspective on RQ3 and RQ4. Using the technique of pattern matching (Yin, 1994), we compared the patterns of the separate cases to each other, and we compared these cross-case patterns to our predictions.

5.1 RQ3: Teachers' Interaction with the Design

5.1.1 RQ3a: Writing Task Design

Teachers were instructed to design reading-writing tasks based on the principles of writing-to-learn. The prompts that teachers formulated are presented in Table 5.2 (for elaborations, see Box 1).

Teachers B and C designed writing tasks that fully met the design principles. With respect to T3, Teacher A adapted the principle of conciseness. This adaptation was a forced adjustment; due to an unexpected shortage during the intervention period, T3 was performed in a test setting. Students received grades for their performance, which represented their final philosophy mark for the school year. The book chapter was handed out in advance: students were encouraged to read the source texts ahead of time independently to prepare for the exam.

Table 5.2. Overview of Writing Tasks

Group	T	Domain	Prompt	No. of sources	Total words (M)
A	T2	Social/political philosophy	Were Leopold and Loeb fully responsible for their actions, or did they have free will from which their actions sprang?	5	825 (165)
	T3		What does decision-making look like, ideally?	1*	8372
B	T2	Philosophical anthropology	To what extent are humans defined by 'reason' when compared to animals?	4	391 (98)
	T3		To what extent should humans allow emotions?	3	410 (137)
C	T2	Philosophical anthropology	To what extent is different treatment of men and women desirable in our society?	4	526 (131)
	T3		Should teachers at your school impose fewer obligations and rely more heavily on students' own responsibility?	4	474 (119)

*A book chapter was used as a source. This chapter contained different philosophers' views on the same topic. The students read the book chapter as a preparation before class, i.e., students had more time to write during class.

Box 1: Teacher-Designed Writing Tasks

Teacher A

Teacher A developed two tasks pertaining to social/political philosophy. For T2, Teacher A wanted students to discuss the issue of "free will". In the introduction to T2, the teacher described the famous case of Leopold and Loeb* (355 words). This description was followed by the following prompt: *Were Leopold and Loeb fully responsible for their actions, or did they have free will from which their actions sprang?* Four source texts were provided, which were all excerpts from the same book. In the first source (178 words), the concept of "insanity" was defined. The second text (403 words) described the history of Richard Kuklinski, who could not avoid becoming a hitman because of his life circumstances. The remaining three texts each defined concepts connected to the issue of whether free will exists: the principle of alternative possibilities (39 words), determinism (94 words), and libertarianism (111 words).

* The names "Leopold and Loeb" refer to Nathan Leopold and Richard Loeb, two wealthy university students who kidnapped and murdered 14-year-old Bobby Franks in Chicago, Illinois, United States, in May 1924. They committed the murder in hopes of demonstrating a superior intellect that entitled them to commit a "perfect crime" without facing any repercussions.

Group A's T3 asked the following question: *What does ideal decision-making look like?* Students were instructed to apply the question to a predefined current situation (note: at the time, Dutch farmers were protesting against the government because of governmental plans regarding the limitation of nitrogen emissions). Students were provided with one source text: a book chapter containing approx. 8300 words. In this chapter, four philosophers (Plato, Dahl, Mill, and Schumpeter) each explained their views on democracy. Students were instructed to refer to at least two of these philosophers' views.

Teacher B

Teacher B designed two tasks pertaining to philosophical anthropology. T2 was titled "the reasonable animal" and focused on the difference between animals and humans with regard to reason. The task asked the following question: *To what extent are humans defined by reason?* Four source texts were provided. The first source (31 words) was a quotation from Aristotle stating that humans are the only creatures gifted with reason. The second source (74 words) was written by the philosopher Immanuel Kant, who argued that humans differ from things (including "mindless" animals) in rank and dignity. In the third source (178 words), Friedrich Nietzsche put man's role on earth into perspective. In the fourth source (98 words), the biologist Frans de Waal discussed why humans are inclined to downplay animals' intelligence.

T3 asked the following question: *To what extent should humans allow emotions?* Three sources were provided. The first source (185 words) was an excerpt from one of Seneca's letters to his friend Lucilius, in which he discussed stoicism. In the second source text (113 words), Aristotle argued that emotions should be allowed, as they can be purifying. In the third text (67 words), the contemporary philosopher Martha Nussbaum claimed that emotions can be a valuable source of knowledge.

Teacher C

Teacher C designed two tasks addressing philosophical anthropology. T2 focused on gender equality. The prompt was as follows: *To what extent is different treatment of men and women desirable in our society?* Four source texts were provided. The first text (157 words) was from philosopher Jean-Paul Sartre, who argued that we have no "inborn essence"; we are free to choose our identity. The second text (96 words) was written by Dick Swaab, a neurobiologist, who argued that men and women are naturally different from each other. The third source (111 words) was from a journalist who wrote a column in a newspaper about her observation that in kindergarten, children are already confronted with typical boys' and girls' toys. The fourth text (116 words) was written by a college student who indicated that medical tests are more focused on male bodies than on female bodies.

T3 focused on autonomy, applied to the students' own school. Students were presented the following statement: *Teachers of [school name] should impose fewer obligations and rely more heavily on students' own responsibility.* Students were then asked to discuss different perspectives and weigh arguments. Four source texts were provided. In the first source text (127 words), philosopher Isaiah Berlin discussed the concepts of "positive" and "negative" freedom. In the second text (94 words), philosopher John Stuart Mill claimed that paternalism is allowed in some cases but not in others. In the third source (89 words), professor of developmental neuroscience Eveline Crone claimed that adolescents' brains are not yet mature. The final source text (100 words) was from Michel Foucault, who argued that freedom is unnecessarily restricted by institutions such as schools.

Overall, the guidelines to make each task performable within a period of 30-40 minutes turned out to be difficult. Teachers experienced difficulty selecting sources and delved into long source texts due to their desire to cover the

whole issue. This difficulty might have been caused by subject-related traditions: teachers of philosophy are accustomed to more extensive writing tasks, in which context students are regularly required to write at their own pace. However, in our study, we asked teachers to design shorter writing tasks because we wanted students to write within sight of their teacher, thereby enabling teachers to observe and support the process. The element of writing in class and within sight of the teacher was valued. Teacher B indicated that this approach was insightful; it made visible the fact that students differ greatly in terms of pace. Moreover, they noted, setting aside time for writing in class makes its importance more obvious.

Both students and teachers thus experienced the tasks as "different", not because of their evaluative nature but rather particularly due to their shortness. We might wonder about the extent to which short writing tasks are suited for the development of philosophical writing. The norm in philosophical writing tasks is that philosophical thinking is central. The thinking process is relevant and reflected in the text. Subject-specific genres, such as philosophical essays, even revolve entirely around thinking processes. In the short tasks used in our study, students seemed to adopt this default view of philosophical tasks: a transfer from one philosophical task to another. Teacher B indicated that if students had viewed the writing tasks as more similar to, for example, test questions, in which context time pressure is common, they would have approached the task in a more pragmatic and result-oriented manner. With respect to the short writing tasks, students did not seem so pragmatic: "Let's write a quick conclusion to it" was not an idea that occurred to them. Nevertheless, teachers valued the short tasks due to their ability to help students develop the skill of writing a philosophical text in an efficient way with the aim of mastering the procedure of reading and writing in the field of philosophy. To reflect the typical thinking and writing procedures used in philosophy more accurately, teachers suggested extending the time available for thinking, for example, by spreading the writing over two lessons or by allowing a second draft to be written after feedback lessons.

Another challenge for teachers was the task of estimating the level of difficulty of primary texts that students could read and interpret individually. Although students were accustomed to discussing primary philosophical texts in class, they usually read those texts collaboratively. Teacher B indicated that they had underestimated how difficult it was for students to read primary texts individually. They wondered whether it would have been more

effective to use longer but more accessible texts and to provide more explanation and decreased information density instead of short but complex excerpts. Although they were not instructed to do so, teachers constructed tasks with increasing levels of difficulty. According to the teachers as well as the jury teams, task difficulty increased from T1 to T3. Difficulty-increasing factors included (1) the complexity of the issue in general (how abstract is the topic, whether the task contains philosophical concepts that are assumed to be familiar), (2) the number and complexity of the source texts (how many perspectives are to be included, whether any new concepts must be understood, whether the student has prior knowledge of the author/philosopher in question), (3) the coherence between the prompt and the sources (how the sources can be related to the issue), and (4) the degree of coherence among the sources (to what extent the sources can be related to each other). Teachers agreed that T1 was of a "basic" level; the issue on which T1 focused was not particularly complex, contained a relatively high number of current sources, was written by "new" philosophers, and included no texts from classic philosophers.

5.1.2 RQ3b: Writing Strategy Instruction

The strategy instruction consisted of three main elements, which are addressed below: the RTW strategy, the video with modeling peers, and feedback lessons.

RTW Strategy. Students received direct instruction by means of a "cheat sheet" that presented the RTW strategy. Drawing attention to students' reading-writing process was a distinctive feature of the instructional design, which was indeed experienced as such by the teachers. Teacher A identified the process instruction using the RTW strategy as an addition to the writing instruction that they would usually provide, which they used to focus mainly on the product: the eventual text. They were familiar with the approach of providing students direct instruction on text structure: for example, by identifying necessary elements in an introduction. Teacher A remarked that they would now know how to provide process instruction in the future because of the didactic tools with which they were provided.

Teacher B also valued the process instruction: when philosophy teachers paid attention to the processes associated with the development of philosophical skills, they thought that this approach would make the subject more "learnable" for students. Teacher B noted philosophy is commonly viewed by students as a subject that is mainly suitable for high achievers. However,

when philosophy teachers highlighted the development of processes, they encouraged students' growth mindset instead, thereby making the subject more attractive.

According to Teacher C, the strategy itself was perceived as "extensive" by students and therefore "impossible to perform during a 30-minute period". However, after the two practice tasks, Teacher C did observe improvement in students' ability to write a good text in 30 minutes. They included a similar writing task in the final exam for the school year, and they were amazed by students' performance. Students wrote up to 600-800 words, which Teacher C thought was "impressive".

Modeling Video. With regard to the video, Teacher B indicated that it was useful with regard to students' self-reflective behavior. As this teacher explained, "The moment they saw other students demonstrating the strategy, they started to empathize. Do I recognize myself in this, do I do this too, (...) and which type of writer do I recognize myself in? It resulted in a good class discussion." Students and other teachers agreed with this claim; however, some students indicated that this situation caused them to feel unsure about their own approach when it deviated from the approaches taken by the students in the video.

Feedback Lessons. The instructional unit contained two feedback lessons (Lesson 2 and 3), each of which followed a writing assignment (T1, T2). In general, teachers exhibited positive attitudes toward the feedback lessons. As Teacher A said, "Students discovered that others could discuss the same issue with a completely different approach. That was very enlightening." Indeed, this enlightenment concerning the available possibilities was one of the goals of presenting the example texts.

Teachers indicated that students were motivated to discuss the texts. The interviews with students confirmed this claim. Overall, students were aware of the main goal underlying the exemplar discussions. As one student (C1) noted, "By reading someone else's text, you sometimes see what you can change about your own text."

However, teachers also experienced difficulties. The main difficulty pertained to the task of directing students toward critical criteria for philosophical texts. Teacher B noted that Group B's list of criteria (which was, as instructed, the result of Lesson 2) was filled with generic criteria for high-quality texts; for example, "the text should contain coherence" and "the issue should

be introduced properly". Nevertheless, Teacher B was satisfied with the result. As this teacher noted, "The criteria list relied more on structure than on philosophical content, but I think that it creates space for students to internalize the content, and then they have more confidence in how they structure their text."

Teacher C also noted that students often suggested generic criteria. They found it difficult to respond to such suggestions because they questioned their relevance for philosophical writing: "Students suggested, for example, that this text has a lively introduction. Yes, okay, it does indeed have that, which is good, but do I really think that is important? It was very new to me, and I didn't really have a picture myself yet; I still don't". This quotation shows that Teacher C would have preferred disciplinary criteria or at least a disciplinary focus for generic criteria; however, this tacit knowledge was difficult to bring to the surface.

Another difficulty for teachers pertained to the need to continue linking the product (what is a good text) to the process (how to write a good text). Teacher C said, "From Lesson 2 onward, we deviated from the approach of how to tackle the task. We got more into the content, of what a good text should actually contain."

Teacher B made an adjustment to the design-as-constructed in Lesson 2, which was a feedback lesson. They considered Group B to be quite a competitive group. Therefore, they were hesitant to select a weak and a strong example text; they thought that doing so would impact students' sense of safety. Instead, they chose to select example texts based on students' writing routines. Teacher B selected one text from a student whom they considered to be a free writer and one text from a student whom they suspected of having engaged in preplanning. During Lesson 2, Teacher B learned that the strategies they suspected the two students of having used were in fact not those that the students claimed to have used: "It turned out to be false, but that did not matter because it was visible that the students wrote from different angles." Teacher B reported that as a result of the adjustment, the students analyzed the texts rather than assessing them, which they believed to be favorable.

5.2 RQ4: Student Progress

During the reflective interviews, all teachers noted that they expected students to have improved their writing skills. Teachers A and C thought that this improvement would be visible in students' texts; however, Teacher B hesitated to endorse the same opinion. Nevertheless, Teacher B thought that

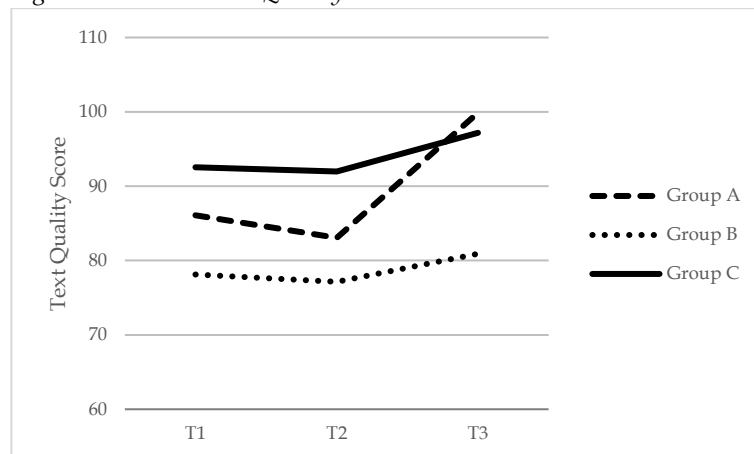
students had improved in the sense that they had gained awareness of their processes.

From the benchmark selection procedure, one main quality criterion emerged: teachers wanted students to demonstrate their independent thinking. Teachers indicated that it was most important for students to relate to the philosophical issue or source texts at hand and to consider different arguments and to develop a personal opinion that could indicate independent thinking. Teachers noted that their selected benchmarks exhibited increasingly independent thinking from T1 to T3. The jury teams agreed with this conclusion. Students' actual scores resulting from the text rating procedure and text length are presented in Table 5.3 and visualized in Figure 5.5.

Table 5.3. Mean Text Quality Scores and Number of Words

	Group	T1 M (SD)	T2 M (SD)	T3 M (SD)	ES T1-T3
Text quality scores	A	86.1 (19.9)	83.1 (19.9)	99.9 (19.9)	.70
	B	78.1 (28.0)	77.1 (29.5)	80.9 (33.3)	.09
	C	92.6 (17.0)	92.0 (18.3)	97.2 (12.0)	.31
Productivity: Number of words	A	336.7 (164.5)	274.9 (73.7)	530.3 (137.8)	1.28
	B	279.4 (87.0)	267.1 (120.2)	287.3 (109.6)	.08
	C	314.6 (71.4)	325.7 (119.1)	415.5 (113.6)	1.06

Figure 5.5. Mean Text Quality Scores at Each Measurement Occasion



At first glance, the flat lines shown in Figure 5.5 for Groups B and C may seem to indicate "no progress". However, the effect sizes for Groups A and C are not trivial: $d = .70$ in Group A and $d = .31$ in Group C (Table 5.3). When compared to the grade effect of a national baseline (Vandermeulen et al., 2023), the effect for Group A approximates a learning effect for writing of one school year ($d = .59$). Moreover, since benchmark text quality increased while the tasks became more complex, a flat line in Figure 6 is likely to indicate improvement in students' philosophical writing.

As noted, Group A's scores increased substantially from T2 to T3. Teacher A thought that the test setting at T3 had positively influenced the results. Students were motivated to write good texts because they aspired to obtain good final grades. In addition, students had time to prepare the writing since they had received the source materials in advance. Teacher A noted that students who had read and analyzed the source materials actively and thoroughly during the class prior to the writing session also wrote better texts. Moreover, this teacher attributed the "drop" at T2 to the difficult class circumstances, which influenced students' motivation negatively.

Overall, Group B scored lower than Groups A and C. The mean scores for this group were approximately 80 points, which was quite far from the benchmark score of 100. Teacher B was surprised by this result, which caused this teacher to consider whether their expectations of students' performance were unrealistic. Group B also wrote the shortest texts on all measurement occasions. To explain this finding, Teacher B's view that philosophical thinking is difficult to exhibit in the context of short tasks might have played a role. As this teacher stated, "I think students did not perform the tasks in a results-oriented way. And I'm afraid I teach this myself, in the sense that I don't pretend that students write such a text in three quarters of an hour impromptu. I think they have to think and weigh and reflect longer." As a result, the texts in Group B were unfinished, and incomplete texts were given low ratings by the jury team. Furthermore, Teacher B indicated that their group was not very high-performing in general but that they had ignored that fact when selecting benchmark texts; their choice for the benchmark texts was based on what they would expect from the "average student" in a regular performing group.

In all three groups, students' actual level was lower than expected by the teachers: teachers had expected that the average student would score 100, which was the level of the benchmark text. However, only Group A was capable of achieving a mean score of 100 at T3, although Group C was close to this mark. Teachers thus consistently overestimated their students. This

discrepancy is in line with previous research on the accuracy of teachers' judgment, which has shown that teachers tend to overestimate their students, particularly low-achieving students (Südkamp et al., 2012).

Markedly, in none of the groups was progress visible from T1 to T2, while strategy instruction was provided between those measurement occasions. A plausible explanation for this finding might be the increase in task difficulty from T1 to T2. T2 and T3 were consistently assessed to be more complex and to contain more philosophical concepts. Moreover, T2 and T3 were embedded in content lessons, thus requiring a philosophical knowledge base. Intervention lessons were therefore interspersed with content lessons.

An alternative explanation for this finding might be that students received feedback only after writing T2. This feedback session was perceived as very instructive; however, any advantages from this session could not be reflected by students' scores for T2.

Productivity was considered to be an additional indicator of text quality. Previous research has shown that good writers write longer texts (Ferrari et al., 1998). In our study, this concept also seemed to be applicable, since text length and text quality scores were correlated: the longer the text, the higher the quality (T1: $r = .607, p < .001$; T2: $r = .793, p < .001$; T3: $r = .684, p < .001$). At T3, Group A wrote the longest texts; however, this major increase in Group A was likely due to the test setting. Moreover, since source materials were read ahead of time, students had more time to write, which is likely to result in longer texts. In Group C, students also wrote longer texts at T3 than at T1, while the length of the texts produced by Group B remained equal. The restricted time might have played a role in this context. As Teacher C noted, "Maybe this is just what a good writer can achieve within the timeframe provided".

We thus conclude that the average student in Groups A and C wrote better texts at T3 than at T1 but that teachers estimated that students' average would be higher.

6. DISCUSSION

With the aim of fostering secondary school students' philosophical writing, we conducted a design study that consisted of three phases: design, implementation, and evaluation. In the first phase, we designed an instructional unit that was open to contextual modifications (RQ1-2). Based on design principles for effective writing tasks, three philosophy teachers (A, B, and C)

designed innovative source-based writing tasks to use in their 10th grade classes. They subsequently implemented the tasks and provided additional discipline-specific writing strategy instruction, which was designed by the research team (RQ2) based on previous writing research. Since it has been reported that teachers' teaching practices are guided by their beliefs and conceptions of teaching, the implementation was prepared and supported by guiding sessions that stimulated philosophy teachers' thinking about writing and writing support in their discipline.

After teachers had implemented the instructional unit in their specific context, we conducted evaluative interviews regarding contextual adaptations. We explored teachers' and students' interaction with the instructional design to test its resilience in open, ecologically valid situations (RQ3).

The results indicated that our design principles for effective writing tasks were feasible. The most challenging aspect of writing task design was to ensure that the tasks remained short. However, teachers did highlight opportunities to improve their tasks, which can be viewed as an indication that the design principles had become successfully internalized at that point. Regarding strategy instruction, the results showed that teachers integrated the design with their contexts within the design parameters and judged the design to be feasible. Teachers indicated that they felt supported in the task of providing process instruction. However, they also experienced that functional knowledge of the criteria associated with philosophical texts was required to provide feedback, a situation that they noted required effort. This aspect remained underemphasized in the teacher guidance program.

Furthermore, we aimed to investigate whether the unit was effective for students' philosophical writing development (RQ4). External jury teams assessed the developed writing tasks and students' texts. We conducted reflective interviews with teachers concerning the expected and actual learner effects.

We conclude that the design-as-constructed showed indications of effectiveness for the development of students' philosophical writing. After the intervention, students in Groups A and C exhibited more independent philosophical thinking in their texts, while tasks were judged to be more complex at T3 than at T1. Student progress was in line with the expectations of Teacher A and C; however, their expected level of student achievement was not fully met; teachers had estimated that students' writing would have improved despite the fact that the task difficulty had increased.

Group B exhibited no increase in writing quality: however, Teacher B reported that students had gained an awareness of their writing process, which they thought was valuable. Moreover, even a flat line from T1 to T3 might have indicated growth since students were able to continue at the same score level even when the task difficulty and benchmark level increased. The concept of progress according to teachers thus involved students' improving ability to complete tasks of increasing complexity.

Since we did not measure teachers' change in beliefs directly, we cannot draw direct conclusions concerning the change in teachers' beliefs regarding writing tasks and instruction. However, indications of change were visible in the interviews since teachers contemplated their role in the context of in-class writing and scaffolding students' individual reading-writing process, thus indicating a student-oriented approach (Kember, 1997).

Four main issues highlighted by this study merit further discussion: teachers' perspectives on what the functions of writing could be (§6.1.1); what a high-quality philosophical text actually entails (§6.1.2); the value of writing strategy instruction (§6.1.3); and professional development with regard to literacy teaching (§6.1.4).

6.1 Main Issues

6.1.1 The Functions of Writing in Philosophy Education

In our study, we instructed teachers to design writing tasks that were suitable for a timeframe of 30-40 minutes. Our rationale for the use of short tasks instead of longer tasks was twofold: (1) we aimed to have students write within sight of the teacher to facilitate process support, and (2) we aimed to enable students to master a procedure and to provide opportunities for them to practice this procedure. In this case, shorter tasks would be most efficient.

Teachers and students experienced the nature of the writing tasks as new: they were used to more extensive writing tasks that featured extended writing time. Although they appreciated the benefit of efficiency with regard to instructing students in the procedures associated with philosophical reading and writing, they also wondered whether the tasks were appropriate to the stimulation of philosophical reasoning. The tradition of philosophical writing advocates "slowing down" and "chewing" on the matter. With short writing tasks, the benefit of "slowing down" diminishes, especially when the tasks are too long to complete within a given timeframe. The balance between the length of the task and the time given should be optimal. It was challenging for teachers to ensure that the size of the tasks remained sufficiently limited to

provide room for discovery on the one hand and to restrain discovery in favor of communicative goals (presenting the key issue in a short text) on the other hand (Baaijen & Galbraith, 2018). However, teachers valued the fact that students wrote within their sight, thus giving them the opportunity to support students' reading and writing process. This fact exposed differences in pace among students that these teachers had never previously considered.

Teachers identified several ways of improving their writing tasks to make them more appropriate to the discipline of philosophy: (1) writing tasks could be shortened (i.e., less sources) to mitigate the philosophical complexity and narrow students' focus solely on key issues; (2) writing could be spread over two lessons to extend the available writing time; or (3) the writing lessons could be viewed as "first draft" lessons, and after a feedback lesson, students could be provided with an opportunity to rewrite their drafts.

6.1.2 Disciplinary Text Quality Criteria

In the instructional unit under study, we included discussions of exemplars to improve students' writing performance in feedback lessons. Since exemplars have proven to be most useful when employed in a dialogic way, with teachers and students jointly establishing a list of criteria for high-quality texts (Carless & Chan, 2017), teachers played an important role in this learning activity. Although both students and teachers valued the discussion of exemplars highly, this component of the strategy instruction raised several issues.

The first issue that teachers highlighted was that students were capable of formulating criteria, but not always the criteria that they considered to be the most relevant. Students, for example, focused on generic criteria for text quality, such as "a good structure", "clear formulations", or "no spelling errors". Teacher-led discussions are thus crucial to guide students away from solely meeting standards in favor of a focus on more discipline-specific criteria, which might be less superficial (Handley & Williams, 2011). For example, teachers considered it to be important that students' texts deliberated on different perspectives instead of merely summarizing them.

This issue of criteria generation remaining at a superficial level (e.g., structural or spelling issues) during class discussions raised another issue: teachers are required to have criteria and standards that are clear to themselves to be able to communicate or discuss these with students. Teachers must have the confidence and functional knowledge necessary to do so. However, this awareness did not appear to be self-evident. Teachers noted that they needed practice to clarify for themselves the requirements that they actually would

prioritize. Knowledge of Goldman's (2016) core concepts of disciplinary literacy would probably be helpful, and this framework should have been included in Session 1.

Teacher C indicated that they felt more confident after several rounds of "practice" discussing exemplars: after the intervention, they felt that they knew the difficulties that students experienced during writing and the mistakes that were commonly made, which made them more confident. Guiding activities (benchmark selection, reflective interview) might also have offered additional value by bringing tacit knowledge to the surface. A related challenge for teachers was to continue establish a connection between product criteria and the process that was the subject of the instruction. However, in this case as well, practice might lead to more routine behavior.

A third issue was raised by Teacher B. They thought that exemplars might be intimidating for some students, constrain their creativity, or result in copying. Are these fears reasonable? Exemplars are not intended to be model texts; exemplars do not demonstrate "how a student should write a text" per se but rather show how a peer approached the same assignment. Exemplars are thus not perfect texts and are therefore unlikely to be intimidating. However, Teacher B's remark raises a question that was also posed by Handley and Williams (2011) regarding whether students consider good exemplars to be models and whether their availability facilitates copying. However, although exemplars might provide students with ideas, these new ideas are intended to expand their repertoire. This fact is not a limitation of creativity, but it contributes to seeing possible ways of elaborating a text and expanding one's linguistic repertoire. Moreover, in our instructional unit, students did not discuss exemplars of topics about which they could write in the future but rather discussed only exemplars of topics about which they had already written. New tasks always contained new topics, which complicated copying behavior. Elements that can be used in new texts might be rhetorical solutions or strategies, which might only enrich students' future writing.

6.1.3 The Value of Writing Strategy Instruction

None of the teachers in our study made adjustments to Lesson 1, which contained direct strategy instruction and modeling peers demonstrating the RTW strategy. One explanation for this fact might be that this lesson offered little room for adjustments. In general, teachers seemed to be satisfied with the strategy and with the concept of process instruction in general.

The question of whether students utilized the strategy or changed their strategy after receiving the instruction remains unclear in this study. In the evaluative interviews, most students reported "no major changes" in their approach. However, aspects of the strategy were noted to have been adopted, for example, ways of integrating sources into a text. Accordingly, the elements that students described as having changed in their approach might be indicators of change even if they themselves mentioned experiencing "no change".

Foremost, the strategy instruction caused students to reflect on their process according to teachers' observations. Monitoring is also part of the strategy itself; at Step 4 of the strategy, students are prompted to keep sight of what is actually asked. In addition to that explicit reference, students' attention was drawn to "monitoring" visually in the "cheat sheet". From this perspective, monitoring might be the most impactful component of the strategy; after all, it characterizes both expert readers and writers (Bråten & Strømsø, 2011; Ferrari et al., 1998). Directing students' awareness toward this aspect of the writing process seemed to be a valuable addition to philosophical writing instruction.

6.1.4 Professional Development

The set of guiding activities formed a distinctive type of professional development (PD) that aimed to help teachers feel equipped to incorporate writing tasks and instruction into their future teaching. With our teacher guidance activities, we met most of the criteria of effective professional development recommended by Desimone (2009); only collective participation was absent. We sought to equip teachers to teach disciplinary writing in three ways. First, we provided them with ready-to-use instructional materials to facilitate strategy instruction to ensure coherence with teachers' prior knowledge. Second, we offered design principles for the creating of writing-to-learn tasks. Teachers thus learned by doing. Third, we organized individual guiding activities to enhance teacher awareness of disciplinary writing instruction and philosophical text quality to promote content focus. The confrontation with students' results that occurred in the reflective interviews led to reflection on the teachers' perceptiveness. Furthermore, the trajectory spanned approximately six months, which was consistent with the criterion of an ideal duration of approximately one semester.

Based on the results, we conclude that the guiding activities as a whole were successful in promoting teacher awareness of disciplinary writing instruction and philosophical text quality. We equipped teachers with

knowledge related to writing tasks and with the didactic tools necessary to address and support writing processes.

The program could be optimized by including functional knowledge of criteria for philosophical texts, for example, by emphasizing the framework developed by Goldman (et al., 2016). What might perhaps improve teachers' literacy practice further is an exchange of thoughts on text quality with their fellow philosophy teachers (cf. Van Drie & Stoel, 2020) after having experienced the corresponding challenges and profits. This approach would include the fifth critical feature of PD. As philosophy teachers are often alone in schools (philosophy is a subject that is mostly taught by one teacher per school), they might benefit from a learning community that focuses on literacy teaching (Desimone, 2009; Van Veen et al., 2012).

6.2 Strengths, Limitations and Directions for Future Research

This research involved a small-scale study with three philosophy teachers with the goal of exploring their interaction with an instructional unit, thereby providing writing process instruction to enhance students' philosophical writing. Our qualitative, contextualized approach provided us with the opportunity to explore this topic in depth. The reflective interview method we used was innovative and insightful. Teachers were asked to reflect on group results, which were produced by an independent jury team, thus ensuring that the student factors that normally influence assessment to mitigated completely. For teachers, this method was highly informative. The confrontation with the group results automatically resulted in reflection.

A fact that must be considered, however, is that the three participating teachers chose to participate in the study and were thus already inclined to improve their teaching and make efforts to accomplish this goal. That fact may well cast the findings of this study in a certain light.

Another limitation might be the procedure we used to measure students' progress, such that writing tasks differed from group to group in terms of measurements, which might be viewed as a threat to internal validity. However, the rationale underlying this procedure was to design an instructional unit that was open to contextual modifications. Providing teachers with the opportunity to design tasks that are tailored to their own curriculum would be a major boost for ecological validity. Furthermore, the writing task design was connected to teachers' expectations of students' performance level, which was within the scope of our study.

The results of our study have several implications for both theoretical understanding and educational practice. This study contributes to our understanding of disciplinary literacy development in an understudied field: the subject of philosophy. For example, this research initiated a discussion about what philosophical writing is, what it could be, which cognitive processes it might or should address, what teachers expect of their students regarding the writing of philosophical texts, and what they consider to constitute student progress.

Furthermore, we provided insights into teachers' handling of innovative learning materials as they enter a new domain at school. These insights revealed that teachers who are unfamiliar with providing writing instruction require guidance and practice in teaching writing processes and discussing text quality with students and that this guidance can be provided by an integrated program of PD activities that stimulate teacher reflection.

Regarding practical implications, we think that our study was significant in two ways. As the philosopher of science De Vries (1984) noted, research can be practically relevant in a technical way by generating new techniques and ways of doing; however, it can also be relevant in a cultural way by generating new ways of expressing, viewing and thinking. Both types of relevance can be observed in this study, as teachers changed their practice and challenged their beliefs and attitudes. As a result, student learning improved. Thus, this study shows that process instruction can be a valuable addition to philosophy education since the results indicated the enhancement of students' philosophical writing.

It would be valuable for future research to explore the extent to which a reflective approach genuinely influences teachers' practice in the long term, as this was not part of the current study. Furthermore, we might explore the effects of intervening with instructional units that focus on philosophical reading-writing processes on a larger scale, possibly by focusing on the effects of such an approach on the most relevant criteria for philosophical writing or by examining this approach in tandem with philosophical reasoning measures.

CHAPTER 6

GENERAL DISCUSSION

The aim of this dissertation was to contribute to the development of secondary school students' disciplinary literacy, with a focus on writing proficiency. Developing students' skills for writing based on sources, which includes reading, is an important educational objective. In their future studies and working life, students need to be able to critically read and assess information, and to communicate their thoughts and reasoning in new texts. Since every discipline knows its own subtleties with regard to reading and writing, it is important for teachers to address these disciplinary nuances. When disciplinary literacy is mastered, a student has achieved the highest level of literacy.

In the Netherlands, which is the setting of our study, the development of students' disciplinary literacy has not been not a priority. Subjects generally center around content goals; communicating knowledge and disciplinary reasoning is underrepresented. However, previous research has proven the potential for teachers to address disciplinary literacy in multiple disciplines for students' literacy levels (De La Paz & Felton, 2010; De La Paz et al., 2014; De La Paz et al., 2017; Hand et al., 2002; Moje et al., 2004).

Students' declining literacy levels, and at the same time, society's elevated threshold, have increased the urgency to respond. In this dissertation, we investigated how teachers can address disciplinary literacy in their classrooms without disrupting their regular teaching. Because the subjects history and philosophy both are known to have high literacy requirements, we decided to focus on these two disciplines. The main research question was:

Which instructional approach is suited to develop students' disciplinary literacy effectively and efficiently in upper secondary history and philosophy education?

We investigated this question through design research: after examining the context and literature, we formulated design principles for writing tasks and strategy instruction. We used the task principles to help teachers develop

sound writing-to-learn tasks. The strategy instruction principles formed the basis for an instructional unit that was tailored to each subject (history and philosophy) and tested in practice.

1. OVERVIEW OF RESULTS

In a prestudy, we explored the context of content teachers' practice around writing task design, writing instruction, and writing support. The goal of this prestudy was to obtain a clear picture of history and philosophy teachers' writing practices and their needs for improving disciplinary writing instruction. Although we started the project with this exploration in an interview study, we did not report on this study with teachers until later, in Chapter 3.

The purpose of the prestudy was to connect with the then-current situation of teachers' teaching and support of writing in history and philosophy, and to disentangle general and discipline-specific processes. In the prestudy, we conducted interviews with teachers of history ($n = 10$) and philosophy ($n = 11$) from different schools. We asked them what type(s) of writing tasks they commonly used in their subject at the upper secondary level. Furthermore, teachers were asked to bring a representative example of a writing task for their subject, and we used this example task as a "hook". We asked the teachers what demands they made on the product of the task, how they supported the students' writing process, and what they thought the students' process should resemble.

The prestudy showed that teachers had difficulty with the verbalization of writing processes. Both philosophy and history teachers expressed that they found it challenging to describe what students should actually do in order to successfully complete their assignment. The majority of history teachers interviewed, disregarded aspects of the writing process; they focused on reading and evaluating sources rather than, for example, on structuring or revising their previously written text. Philosophy teachers were more aware of the role of writing processes in their assignments, implying that they recognized writing as a means of learning philosophy.

The first study reported on in this thesis (Chapter 2) was a background study of students' processes while performing writing tasks for the two concerning subjects in a think-aloud study. Fifteen high-achieving students were selected, since we aimed to explore which cognitive processes novice-experts would

show when they were reading and writing in the two disciplines. Our research questions were:

RQ1: Which patterns in source-based reading and writing can be related to (a) the quality of text produced by students, and (b) the quality of students' thought processes?

RQ2: Are there differences with regard to RQ1 depending on the discipline under consideration (history versus philosophy)?

For each subject, we selected a writing task based on sources, which was thought to be representative of the subjects in question, based on the interview study with teachers. For history, the task presented to students contained a source text, of which students had to write down whether it was useful to be able to answer a certain research question. For philosophy, the task required students to reflect on a specific statement. Students of 11th grade who took both subjects, performed the two tasks in random order, while thinking-aloud.

To identify students' cognitive processes, we employed a data-driven approach to find patterns in students' processes. For the segmenting and coding of students' think-aloud protocols, we also made use of existing frameworks for reading and writing in general (e.g., Hayes & Flower, 1981), and reading and writing in the disciplines (e.g. Brante & Strømsø, 2018; Corcelles & Castelló, 2015; Hof et al., 2015). Three main categories: reading, writing, and metacognition, were subdivided into 11 process activities. All process activities were analyzed on frequency, and absolute and relative duration.

Subsequently, students' protocols were assessed on two measures of quality: (1) quality of the eventual text, and (2) quality of students' thought process. Afterwards, quantitative analyses were conducted: we researched correlations between the quality of students' texts, the quality of students' thought process, and their process activities.

Results showed that the relations between quality and activities differed per task, or, as we suggested, per subject. In the history assignment, text quality mostly depended on planning variables. Students who spent more time on planning and planned more frequently, produced higher quality texts. By contrast, in the philosophy assignment, most of the variables affecting quality were writing variables. Students who wrote relatively high-quality texts

showed characteristics of expert writers: they produced longer texts, spent more time writing, and reviewed more often.

Furthermore, the relation between text quality and thought process quality differed per subject. The history assignment mainly provoked the outcome of students' reasoning process, while the philosophy assignment invited students to write down their entire reasoning as well; the tasks stimulated thinking and writing to codevelop.

The results of this contextual study had implications for the next steps in our research: we considered the character of the philosophy assignment, which evidently could stimulate students' thought process, to be a suited task type for use in other content classrooms as well. As opposed to the history assignment, which only required an outcome, the philosophy assignment could be characterized as a writing-to-learn task; such tasks are well suited for use in content classrooms, since they fit the focus on knowledge development.

In the next study (Chapter 3), design principles and an instructional design to enhance students' historical writing were developed. Furthermore, the instructional unit was evaluated on validity, practicality, and effectiveness. The main research question was:

RQ: What are design principles for a valid, practical and effective design for writing tasks, and writing instructions, for upper secondary school history to support students' historical writing?

Insights from our prestudy, think-aloud study (Chapter 2), and a literature search, led us to the formulation of two design principles, which formed the foundation of the instructional units for both history and philosophy. The two design principles were:

If we want students to develop a profound understanding of content through writing, then:

Design principle #1: students should write short evaluative texts, based on multiple primary sources that represent multiple perspectives.

Design principle #2: students should be provided with discipline-specific, dual-route, reading-writing strategy instruction which is easily applicable for teachers.

The strategy instruction was included to support the construction of coherent knowledge; making the writing process manageable, is likely to promote knowledge construction, and having a dual route to accommodate different writing preferences prevents the writing process from becoming a burden.

To put design principle 2 into practice, a strategy was designed, based on the insights obtained in our think-aloud study, and previous interventions studies, using writing strategy instruction in the content classroom (e.g., De La Paz et al., 2017). This resulted in the seven-step Read-Think-Write Strategy (RTW strategy), which was accustomed to historical evaluative questions, in the form of "to what extent"-questions: questions that were assumed to be writing-to-learn questions. The strategy offered two routes to accommodate different writing process preferences: "free writers" and "pre-planners".

In a trial study, the instructional unit was implemented and evaluated by two history teachers in a switching replications design with three measurement occasions. As measurements, three different writing tasks were designed in cooperation with the participating teachers. This procedure enhanced feasibility and encouraged teachers' engagement with the materials.

Students' texts were assessed by a jury of history teachers on three criteria: holistic quality, content quality, and quality of structure. Results showed indications of effectiveness on content quality. However, results were not fully convincing, possibly because the second principle did not stand out sufficiently. The participating teachers valued the writing tasks as described in principle 1. They doubted however some aspects of the strategy instruction as mentioned in principle 2. The main recommendation arising from this trial was to increase teachers' background knowledge of the design principles, and to explain the underpinnings behind these principles more prominently. In a redesign, a teacher development session was thus included.

In the third study (chapter 4), we tested the effectiveness of the optimized instructional design in a quasi-experimental study with three research conditions, to obtain insights into effects of two factors: (1) implementation of writing tasks, and (2) implementation of writing process instruction. Research questions were:

RQ1: To what extent does the replacement of workbook assignments by evaluative writing tasks contribute to (a) disciplinary source-based

writing proficiency, (b) epistemic experience of writing, and (c) students' course content knowledge?

RQ2: To what extent does the addition of strategy instruction to such writing tasks contribute to (a) disciplinary source-based writing proficiency, (b) epistemic experience of writing, and (c) students' course content knowledge?

RQ3: To what extent are effects on RQ1 and 2 moderated by learner characteristics (writing beliefs and self-efficacy)?

In a first experimental condition (WT+S; 7 groups, 119 students, 6 teachers), participating teachers designed writing tasks, which were in line with design principle 1. Additionally, they implemented an optimized version of the designed writing strategy instruction, using the RTW strategy. In a second experimental condition (WT; 3 groups, 63 students, 2 teachers), teachers also designed writing tasks, but the strategy instruction was omitted. A third condition was a non-writing control condition (C; 4 groups, 86 students, 3 teachers).

In a quasi-experimental pretest-posttest design, with two writing tasks as a pretest and posttest, we aimed to measure effects on students' writing proficiency. These writing tasks were evaluative tasks on a topic which was not discussed during the history lessons. After the strategy instruction - containing direct instruction, through a video showing modeling peers, and discussions of exemplars - students practiced writing through performance of two writing tasks, which were designed by their own teacher, and therefore fully tailored to the content of their history lessons. These writing practice tasks thus differed from group to group.

Students' pre- and posttest texts were assessed by a jury team of history teachers, on five aspects: situational understanding, multiperspectivity, source use, argumentation, and structure. Furthermore, texts were assessed holistically, by means of a text scale. These scores were indicators for students' writing proficiency in history.

Furthermore, we explored the effects of the writing tasks and strategy instruction on students' epistemic experience: to assess to what extent students experienced writing tasks as learning tasks. Finally, we explored whether writing tasks were a proper alternative to other, more common, learning activities for learning history. Therefore, we measured students' course content

knowledge in a pre-post recall test. The topic of this recall test differed from group to group, as each group followed their regular year plan.

Results showed effects on students' writing proficiency: students in Condition WT+S, who had received writing strategy instruction, outperformed other students: they wrote better texts. This applied to students with all self-efficacy levels and beliefs types. Significant effects of the additional strategy instruction were visible in all aspects of text quality, holistically and on discipline-specific criteria. Next to that, students in Condition WT+S more often indicated they experienced the epistemic function of writing, compared to the control condition.

Recall test results showed that students in Condition WT+S performed equally well as students in the other conditions regarding course content knowledge gain: time had been devoted to writing strategy instruction, and was not wasted, as had been feared by content teachers. These results imply that writing seems to be a promising learning activity for history, however, only when attention is paid to the writing process, this results in more advanced writing proficiency.

The second and third study (Chapters 3 and 4) were focused on the subject history. However, to compare different disciplines, we have investigated the effects of a similar instructional design for philosophy in a fourth study (Chapter 5). This study focused on the interaction of three philosophy teachers with writing task design and the instructional unit, which they implemented into their philosophy lessons in 10th-grade. The instructional design was open to context modifications. Implementation was accompanied by teacher guidance activities, utilizing implementation of the intervention as a learning trajectory for teachers. Six activities integrated in the research process were aimed to prompt teachers' contemplation of writing instruction and writing quality. Our main research questions were:

RQ1: Which design principles can be derived from previous research on developing secondary school students' philosophical literacy?

RQ2: How can these design principles be translated into an instructional unit?

RQ3: How do philosophy teachers interact with (a) principle-based writing task design, and (b) an instructional unit to support students' disciplinary writing?

RQ4: To what extent has the instructional unit contributed to (philosophy teachers' notions of) progress in students' philosophical writing development?

This study can be characterized as a design study, to profoundly explore how philosophy teachers interacted with the designed instructional unit, what results it yielded for students' disciplinary writing proficiency, and to what extent students' actual achievement was consistent with teachers' expectations.

In this study, we adapted our instructional unit for history to a next knowledge domain: philosophy. The two design principles for writing tasks and instruction remained similar. The main difference between the two units concerned the RTW strategy, which was tailored to "philosophical questions". The four main aspects (reading, thinking, writing, and monitoring) remained, but contained typical philosophical "colors": students were prompted to contemplate different perspectives in order to establish their own stance towards the issue, to show their independent philosophical thinking, to write from their own perspective, to use philosophical concepts, and to include examples from their own life or current events as illustrations. The accompanying strategy instruction was optimized, but did not change conceptually.

Three philosophy teachers were instructed to design their own writing tasks for use in their 10th grade classes, and implemented the designed instructional unit. Implementation was accompanied by teacher guidance activities, which were aimed to prompt teachers' contemplation of writing instruction and writing quality. We monitored and evaluated teachers' implementation, to explore how resilient the design would be in open, ecologically valid contexts (RQ3) and conducted evaluative interviews with students and teachers to research their adjustments to the design.

Furthermore, we explored student progress as perceived by teachers (RQ4). Students' writing proficiency was measured with writing tasks (T1-T2-T3), two of which were designed by the group's teachers (T2-T3), and thus differing from group to group. Teachers then selected a benchmark text for each task: the "average" text for that task. This activity elicited a reflection on quality criteria in philosophical writing.

To frame teachers' perceptions of progress, the writing tasks and students' texts written in response to these tasks were assessed by independent jury teams, with benchmark texts as references. In reflective interviews teachers explained and reflected on their students' results in reflective interviews. Results indicated that philosophy teachers were able to design qualified writing-to-learn tasks, when instructed on theoretically grounded design principles. The accompanying instructional unit turned out to be feasible and instructive. With regard to students' progress, we carefully concluded that the instructional unit was a valuable contribution to students' philosophical writing proficiency. Students showed more independent philosophical thinking in their text, whilst tasks became more complex. The teacher guidance program elicited reflection on the value of writing tasks and instruction for subject goals, and on perceptions of student progress.

In general, the results of Chapter 5 provoke discussion about what writing tasks and writing instruction can provide for philosophy education and what a high-quality philosophical text precisely entails.

2. CONCEPTUAL ISSUES

The main objective of this thesis was to enhance secondary school students' disciplinary literacy, by improving teachers' practice with regard to writing task design and teaching writing in their respective subject. In retrospect, three lines of development have become apparent through the four studies of this dissertation, namely, with regard to (1) disciplinary writing; (2) supporting writing; and (3) innovation of teachers' writing practice.

2.1 Disciplinary Writing

Before we can think about how to illuminate the disciplinary aspects of writing in subject lessons, we will need to consider the writing tasks employed in content lessons. The think-aloud study of Chapter 2 revealed that writing tasks can highly direct students' cognitive processes. The history assignment used in this study predominantly elicited students' reasoning outcome. Resulting texts, therefore, offered limited insight into students' reasoning underlying that outcome, whereas that might be precisely what is most valuable to teachers.

While we cannot quite say that all history assignments are of such a nature, the one used in the explorative study was selected to be representative of the subject, based on the prestudy. History teachers tend to align their

assignments with what is asked in the final exam: in the Netherlands, that is, answering shorter questions. In Dutch history education, therefore, writing tasks rarely focus on communicating the underlying thought process, however, this does not mean that no reasoning process should take place, or that the reasoning process is not considered important. On the contrary: history teachers put much effort into students' historical reasoning skills. Why, then, would a content teacher not employ a writing assignment to be a vehicle for teaching this reasoning skill as well? Not only will it teach students to express and organize their thoughts on paper, but it will also help the teacher gain insight into the student's reasoning level. A good example is the document-based question (DBQ), an exam task in the US, which might support learning about history as an inquiry into the past (De La Paz, 2011; Sendur, 2021). Therefore, we used tasks similar to DBQ essay tasks in the continuation of our studies. History teachers' evaluations after implementation of the instructional unit (Chapters 3 and 4) showed that they valued writing tasks, when these were aimed to exploit the epistemic function of writing. They seemed uninformed regarding the effects a writing task might have on learning, and their writing task design relied on assessment, rather than on a contemplation of learning effects.

For philosophy teachers, this was much less of an issue. The most frequently mentioned writing tasks were philosophical essays or thought experiments. These are genres that are generally well suited as teaching tools; these genres invite thoughts to be further developed while writing.

All in all, we can conclude that an appropriate writing task for both of these school subjects focuses on learning, which might be interpreted as content learning and/or learning to reason (Galbraith, 2015). Learning is generally enhanced when students are prompted to explore different perspectives on an issue, by means of multiple sources to analyze and interpret, and to provide their own perspectives on the issue based on the sources, in an evaluative text. Since the label "evaluative text" is not quite common in either subject, we translated the genre into "to what extent-questions" for history, and "philosophical questions" for philosophy, which are familiar labels within each subject.

When teachers are instructed on design principles for disciplinary writing tasks, our studies show that teachers are capable of designing qualified tasks. Overall, history teachers were comfortable with the design principles, which suited their subject goals of sourcing and discussing historical issues. Their

main concern was to keep the tasks concise, to keep the number of sources manageable.

Philosophy teachers struggled with the length of the writing tasks for different reasons. They generally recognized the usefulness of short tasks to make students' processes visible in class, but at its core, the task did not correspond to their view that philosophical thinking, and with that, reflecting and writing, simply takes more time than is available in a single class period. Intrinsic to philosophy is the contemplation of an issue; a process that takes time. The prescribed short tasks ran counter to this premise, as the thought process was delineated by the short time. This design principle thus tore into teachers' conceptions of the learning effects of writing assignments. Nevertheless, teachers recognized after the intervention that short tasks can be a different way of utilizing writing for learning, namely, to master the procedure of philosophical writing.

All in all, with this dissertation, we advocate the integration of writing tasks which can promote learning, in order to develop students' disciplinary literacy. In our studies, we used prompts which covered the core of a topic to be learned, as a replacement of workbook assignments, in which core topics are often broken down into a series of shorter questions. Results of our studies showed that teachers valued the writing tasks, history teachers in particular. Furthermore, our experimental study of Chapter 4 showed that students who performed writing tasks and received strategy instruction, experienced the epistemic function of writing. Presumably students are learning, when they perceive the task as relevant.

Including evaluative writing tasks might be a vital counterbalance to the contentious mile-wide-inch-deep approach in history education, where teachers are alleged to cover a large set of topics, but on a somewhat shallow level. Since writing is time-consuming and cognitively demanding for students, we argue to include writing tasks every once in a while, next to other useful learning activities which also have the power to develop students' knowledge accretion and disciplinary reasoning.

2.2 Supporting Writing in the Content Classroom

When writing is used as a learning tool (writing-to-learn) as well as a means to work on literacy (learning-to-write), it is advised to provide strategy instruction. Our studies have shown that disciplines differ: writing processes as well as texts have discipline-specific elements, which should be addresses in

content classes. With our studies, we provide examples of reading-writing strategy instruction implementation in two subjects. The strategy instruction we developed was highly discipline-specific, based on previous research on reading and writing in the specific subjects, combining theoretical understanding of generic reading-writing development, and disciplinary literacy.

Our prestudy (Chapter 3 and 5) revealed several obstacles to overcome. Most importantly, the interviews with history and philosophy teachers showed that, overall, teachers have limited knowledge about writing processes. Moreover, in general, history teachers remain staunch supporters of "content first". This observation raises the question also discussed by Moje (2008), what might be the root cause for history teachers to be so adamant in their adherence to a traditional teaching paradigm. While there seems to be an increasing focus on historical reasoning skills in history education, literacy skills hardly play a role.

An appropriate example of a tenacious teacher is one of the teachers who participated in the history trial study (Chapter 3). They supported the implementation of writing tasks that focused on learning; they called it "the essence of history", to work with evaluative tasks. But for providing writing instruction, to support students in completing such a task, they felt less responsibility. The steps in the strategy that dealt with reading and reviewing sources felt like a logical addition in their history teaching, since these steps addressed typical historical reasoning skills, but that sense of responsibility seemed to disappear at the steps where students transitioned into communicating their reasoning process.

An explanation for teachers' insistence on knowledge accretion might be an overloaded curriculum. Teachers commonly consider a knowledge base to be necessary to develop reasoning skills. A logical consequence would then be for teachers to build students' knowledge first, and subsequently address reasoning skills. Literacy then often remains elusive, due to the time pressure. However, in our study, we tried to consolidate knowledge building, reasoning skill development, and literacy proficiency, as we assume these might develop hand-in-hand.

As in their attitude towards writing tasks, participating philosophy teachers differed from history teachers with respect to their beliefs about providing writing instruction and support. Philosophy teachers already frequently used tasks that could stimulate thinking, and these teachers generally felt responsible for guiding the process precisely, because they aimed thought processes and writing to converge. The main obstacle for philosophy teachers was their

lack of knowledge about reading and writing processes. However, with our guidance, participating teachers rapidly felt familiar with providing process instruction.

A key question in our research, then, was: how can we achieve teacher engagement to ensure that they integrate appropriate writing tasks as opportunities to develop disciplinary literacy, and that they incorporate time for additional reading-writing instruction in their subject lessons? We addressed this conundrum by inviting teachers to develop their own writing tasks to be embedded conveniently in the curriculum. Writing in-class enables teachers to observe and support students' process. Our advice was to make use of minimally invasive short tasks, that take approximately one lesson period to complete. In addition, we developed feasible materials, such as a video with modeling peers, to relieve teachers of the complex, knowledge-requiring components of writing instruction. Finally, we examined whether writing instruction affected content learning (Chapter 4) as a negative effect, a common fear (primarily among history teachers) which is generally brought up as an excuse to avoid writing instruction. Students' recall tests showed that writing tasks, when accompanied by process instruction, were able to compete with regular history education as regards history knowledge acquisition; a reassuring result.

Despite our efforts for a feasible instructional design, the unit presented challenges. The main challenge for teachers was the discussion of exemplars. To address subject specificity in a writing task and text, a teacher will need to be aware of what a good historical or philosophical text actually entails. In practice, teachers are not overly concerned with differences between subjects; they are primarily focused on their own subject. Teachers would like to make use of general writing skills, but do not link general writing skills to subject-specific writing skills because they are not consciously engaged in this matter. Knowledge of subject-specific writing proficiency therefore often remains implicit among students.

This may be, because teachers themselves find it difficult to make explicit what the criteria are for texts in their subject. When exemplars were discussed in class, students often suggested requirements related to general text quality, rather than subject-specific text quality. Generic elements, such as overall text structure, also provide guidance for subject-specific reasoning to emerge; therefore, it is not trivial. Although teachers indicated they were satisfied with generic requirements to some extent, nonetheless a sense of dissatisfaction

prevailed. Of course, both a history text and a philosophy text should have a clear text structure, but what distinguishes an excellent text from a mediocre text in history or philosophy? In fact, the discussion of text quality lingered at the level of basic, or perhaps intermediate, literacy but did not reach a profound discussion at the disciplinary level (Shanahan & Shanahan, 2008). Therefore, it might be valuable if teachers from different disciplines exchanged more thoughts with each other about what they consider constitutes a good text in their discipline. It could make teachers aware of the text and process elements that distinguish their subject from other subjects.

Our study revealed such nuances for the subjects history and philosophy. In response to a "to what extent"-question in history, students are required to take a stance towards the issue, based on a profound investigation of historical sources. Students are required to explore a variety of arguments, to support them with sources, and balance them. "Sourcing" is an important step in this process (Wineburg, 1991; Van Drie & Van Boxtel, 2008). The student's stance towards the issue is explicitly based on the historical inquiry; personal preference has no place. This way, a student will show their capacity to reason historically.

This is different in a student's response to a philosophical question, where personal notes are appreciated, provided that they are supported. In a philosophical text, the student as a writer is visible throughout the text, the text is obviously "colored" by the student's perspective. This personal perspective is what makes the philosophical text interesting: it shows their independent philosophical thinking. Thus, how an answer to an evaluative question should be framed, depends on the discipline to a large extent.

2.3 Innovating Practice

Previous literacy researchers (e.g., Gillis, 2014) have advised to not adopt reading and writing strategies in the content classroom, but to adapt those strategies, and to always keep track of content goals. To meet this requirement, interventions aimed at students' literacy development would need a large amount of flexibility. Over the three intervention studies, therefore, we might recognize an increase of openness of the designs-as-constructed. In the first trial study (Chapter 3), we started with a semi-open design, with two teachers from the same school co-designing writing tasks to use in their classes. In the experimental study (Chapter 4), teachers from multiple schools were involved, which resulted in many different tasks covering a wide range of topics. Moreover, we presented different methods for teachers to choose

from to discuss exemplars with teachers in class. Flexibility of the design was expanded again in the philosophy study (Chapter 5), with highly contextual measurements and therefore an also highly contextual assessment procedure to determine student progress. Although these flexible research designs presented us with methodological challenges (as discussed in the next section), evolving our research design to include more openness in each subsequent study may indicate the need for flexibility in innovating teachers' writing practices. When a design is too rigid, teachers might see too many barriers that prevent them from integrating writing instruction after all.

In our research, it gradually became clear that we needed to guide teachers in implementing writing instruction and supporting students in their writing process. Even though we aimed to keep the interventions accessible, in our intervention studies (Chapters 3, 4, and 5), guidance to teachers intensified and became more reflective. We aimed to add practical significance in the long run; therefore, we aimed to not only enhance teachers' practice, but also challenge their beliefs about writing and writing instruction. We thus aimed for cultural change in addition to technical change (De Vries, 1984).

Eventually, the technical guidance focused on a number of knowledge components: (1) knowledge about writing tasks, which type of task is appropriate when; (2) knowledge about text quality, especially disciplinary aspects; (3) knowledge about (disciplinary) writing processes; and (4) procedural knowledge about how to guide writing processes, and (5) how to discuss writing products. These components will be useful for future professional development programs aimed at teachers' improvement of literacy teaching.

3. METHODOLOGICAL ISSUES

In this section we address six issues concerning research methodology: (1) the power of mixed methods design research; (2) validity; (3) teacher involvement; (4) measuring content knowledge; (5) epistemic experience; and (6) measuring writing proficiency.

3.1 The Power of Mixed Methods Design Research

In this thesis, Chapters 3 and 5 are explicitly presented as design research, aimed at the design of an instructional unit to promote students' historical and philosophical literacy. However, this thesis as a whole can be characterized as educational design research as well: the four studies together represent the different stages of analysis and exploration (Chapters 2 and 3), design and

construction (Chapters 3 and 5), and evaluation and reflection (Chapters 4 and 5).

Next to quantitative measures, we frequently made use of qualitative approaches to obtain more insights in students' processes and progress, which brought many advantages. First of all, we used qualitative methods in the exploratory phase of the study. We conducted interviews with teachers to establish a window into the context in which we were operating. The explorative study of students' cognitive processes while performing writing tasks for two different subjects, was also of qualitative nature. Of course, the think-aloud method also has disadvantages: for example, students' verbalizations are not conclusive of all students' thoughts (Pressley & Afflerbach, 1995; Van Someren et al., 1994). Nevertheless, the think-aloud method provided us with a solid insight into the activities students perform while making writing assignments for history and philosophy. This insight provided a solid base for designing a reading-writing strategy for each subject.

Invariably, we also used a qualitative approach in evaluating the implementation of the lesson series (Chapters 3, 4, and 5). After implementation, we consistently conducted interviews with teachers and students, not only to determine implementation fidelity, but also to get an idea of teacher and student reception. To what extent did the new writing tasks and writing instruction align with teachers' views of their subject, and why? And how did students experience learning through writing?

A drawback of our qualitative approach in the evaluations with students, was that not all students were consulted. Students participated in the interviews voluntarily, which might have resulted in a selection bias. Students who were willing to participate, might have been those who were more motivated than others. We tried to overcome this by asking teachers to encourage a diverse group of students for participation. However, a quantitative measure (e.g., questionnaire) could have provided valuable additional information.

A final beneficial qualitative addition in our research, were the retrospective interviews with philosophy teachers, after the implementation of the design as well as after the assessment of the student texts. To enable the latter, students' texts were assessed by an independent jury. This jury made use of benchmark texts, which were selected per task by the philosophy teachers to represent "the average student's expected performance". The remaining texts written in response of the respective task were assessed in relation to this

benchmark. After the rating procedure, the teacher was presented the jury's scores to reflect on.

In their reflections, teachers not only contemplated the implementation of the instructional unit, but also the level of the writing tasks they had developed, the benchmark texts they had selected, and the scores the students in their group obtained in relation to those benchmark texts. Among other things, this method provided us insights into teachers' views of what exactly constitutes a good text in philosophy. Furthermore, we obtained insights in teachers' perceptiveness of students' writing proficiency level. The fact that the texts were assessed by an independent jury added an extra dimension: it offered us insight into the teacher's assessment of the difficulty of self-composed writing tasks and the teacher's expectations of students' performance level. The in-depth approach in Chapter 5 thus provided us with a flood of observations.

In addition, during the different stages of the research project, we made use of quantitative methods, to measure if students' writing proficiency had progressed after intervening with a writing strategy instruction. In three studies, students' texts were assessed by jury teams (Chapters 3, 4 and 5), and their scores were analyzed using multilevel statistical analyses (Chapters 3 and 4). The larger scale of the Chapter 4 study, with 268 students participating, provided us with evidence of students' progress in their disciplinary writing. Precisely the combination of different methods thus offered great insight, illuminating different perspectives.

3.2 Validity

Generalizability was an important concern for the impact studies in our research project. We wanted to design an instructional unit that could contribute to students' subject-specific writing skills. Generalization across disciplines seemed not to be possible here - after all, it is specifically discipline-specificity that matters - but we did aim to generalize across teachers and students, per discipline. In particular, we aimed for generalizability in a number of areas.

First, we prioritized ecological validity. Our research was conducted in existing groups of students, taught by their regular teacher. We preferred a teaching situation representative of the real world, rather than a researcher in front of the class. Moreover, our aim was to investigate to what extent the subject teachers themselves could cope with the developed writing

instruction. After all, the principles behind the instruction had already proven their value in previous research.

Also, considering ecological validity, and then mainly the integration of instructional materials into regular curricula, we did not completely seal off the lessons: we entrusted the design of writing tasks to teachers themselves, accompanied by a PD session on design principles for effective writing-to-learn tasks. Hereby, we aimed to promote engagement, and to effectuate a knowledge base of teaching writing and writing task development. This decision may have led to create some muddling, as groups became less directly comparable. However, variety existed in all conditions (Chapter 4). We presume generalizability was actually also enhanced by using such a variety of writing tasks, covering multiple topics, in both subjects. Moreover, writing tasks were tuned to different grades, since in the history studies, students of multiple grades (10th-11th-12th) participated.

We also tried to enhance generalizability by consistently exploiting design principles. After implementation of a unit, we evaluated to what extent the design principles behind the design had been brought to bear. Responding to the appeal by Rijlaarsdam and colleagues (2017), to consider an intervention as a construct, and to define and operationalize it as such, we explained our subject-specific interventions as comprehensively as possible in Chapter 3 (for history) and in Chapter 5 (for philosophy). The importance of design principles for implementation also became evident after the trial study of Chapter 3, where we concluded that teachers must be exceedingly aware of the underlying principles, in order to properly implement an intervention.

In addition, the methodology itself enhanced generalizability. First, generalizability was sought during the data preparation phase (Chapter 4), by including a variety of raters in the text assessment procedure. Second, we used statistical analyses, more specifically multilevel modeling, which allowed us to generalize across tasks, students, groups, and teachers.

We might criticize the internal validity of the study in Chapter 4, in which the teachers themselves chose the research condition, rather than being randomly assigned to it. Since we considered motivation of the teacher to be an important factor for successful educational innovation, teachers were offered a choice with regard to the two experimental writing conditions. As a result, we do not know to what extent the positive results could also be due to the teachers' positive stance towards strategy instruction from the beginning.

One might also comment on the validity of the research instruments: although randomization is not always necessary to detect effects, all of our

studies had fixed pre- and post-measurements for writing proficiency. As a result, differences in pre- and post-measurement scores could be due to task differences. For example, prior knowledge about a particular topic could have played a role. When students have a high level of prior knowledge about a topic, they will write better texts (Olinghouse et al., 2015). Therefore, in interpreting the results, we only considered the differences on the post-measurement; the pre-measurement was included as a covariate. However, this raises a next validity issue: in the study of Chapter 4, we only used one task as a post-measurement. We might have to consider the extent to which this task was representative.

However, we tried to avoid any of these biases by asking control questions. We did find differences between the pre- and post-measurements, but they were present in all conditions. In addition, we used a transfer task, to demonstrate that students were able to perform new tasks on a new topic. In the study of the effects of the lesson series on writing skills in philosophy (Chapter 5), we controlled for task differences in a different way: all writing tasks were analyzed in great detail by an independent jury team.

A final comment regarding external validity, is a comment about the limited research group: all participating groups received education in upper secondary levels of the pre-university track. The target group of our study was upper secondary education, because in these grades, students are expected to read and write longer and more complex, abstract texts, which also explicitly address disciplinary literacy. We have not tested whether the materials are also suitable for lower grades or tracks. However, writing based on sources can be done at many levels, and experience shows us that adapting writing tasks to the level of the target group in question is possible. The teachers in our study distinguished between tasks for 10th, 11th, or 12th grade students, seemingly without much effort. Whether strategy instruction in modified form will also be effective, is beyond the scope of our current study, but previous research has proven strategy instruction to be successful in earlier grades (De La Paz & Graham, 2002), and for students with mixed abilities (De La Paz, 2005; Wissinger et al., 2018).

3.3 Teacher Involvement

As mentioned in the previous section, we entrusted the design of the writing tasks to the content teachers participating in our studies. We aimed to involve history and philosophy teachers in the design process for several reasons. First of all, we expected that teacher involvement would make the design

more feasible, and it would enhance generalizability. After all, teachers would not be bound to one topic, nor would the results of the studies. Moreover, the writing tasks would be better writing-to-learn tasks, when these would actually be tailored to the content learning goals as envisioned by the teacher and/or prescribed by the curriculum. Next to that, discipline-specificity would be ensured, and the content teacher would be valued as the expert in their field. Finally, with the involvement of teachers we aimed to effectuate their engagement to disciplinary literacy.

One drawback of teacher involvement in the design process was that tasks not only differed in content, but also in length, complexity, and abstraction. For example, a prompt about a more complex issue required more source texts. As a result, there was a wide variety of tasks, and since our research team mainly holds a linguistic perspective, it was not easy to estimate the extent to which the tasks were still comparable in, for example, the degree of prior knowledge required for the task. At the same time, an added benefit was that we could obtain insights into teachers' interaction with design principles, and which feedback and guidance they required.

3.4 Measuring Content Knowledge Learning

This research project was primarily concerned with promoting discipline-specific writing, rather than content learning. Learning indicators were mainly investigated (in Chapter 4) to explore whether we can promote writing development without impeding content matter. Our main aim was to enhance students' disciplinary literacy, and to gain power in convincing teachers to implement interventions with this aim, we investigated learning effects next to effects on students' writing proficiency.

Foremost, we measured students' learning by assessing students' texts. However, to what extent can learning be measured with writing tasks? What kind of knowledge is enhanced with writing in the disciplines, with the writing genres we used? Philosophical reasoning, one of the main goals of philosophy education, was considered well-represented in resulting texts. Students' level of historical reasoning can also be derived from their texts; however, history seems to depend more on facts and functional knowledge, than philosophy. Therefore, we included an additional learning measure in the Chapter 4 study: a recall test, measuring content knowledge of the topic discussed in class during the intervention period.

Nevertheless, we could question whether this recall measurement was the best instrument to measure content knowledge. Previous research has

criticized recall tests as a measure of knowledge gain through writing, since a recall test is likely to detect knowledge accretion, which is, according to Schumacher and Nash (1991), an "unlikely outcome of writing" (p. 92). Despite this objection, the recall measure demonstrated that knowledge accumulation was not limited by the writing strategy instruction.

3.5 Epistemic Experience

In Chapter 4, we added a third indicator of content knowledge: epistemic experience. This concept refers to the degree to which students experienced the epistemic function of writing: the awareness of constructing knowledge through writing.

The concept was derived from Bereiter's (1980) model of skill systems in writing development. This model includes six stages of development in the maturation of writing, with epistemic writing as the highest stage. Epistemic writing transcends the skill of unified writing: the ability to understand not only other people's perspectives on the text one is writing (what we call communicative writing), but also one's own perspective as a reader of one's own text. As a result, students become more authentic writers, and develop more personal styles.

According to Bereiter's model, epistemic writing emerges when the skill of reflective thought is developed. Creating a text then becomes a personal search for meaning. Writing becomes "an integral part of thinking" (Bereiter, 1980, p. 88). Philosophical essay tasks and evaluative questions based on historical sources rely heavily on unified and epistemic writing, the complex stages of writing.

In the study in Chapter 4, epistemic experience was measured to indicate students' knowledge development. We assumed that the sense of learning contributed to actual learning. Epistemic experience was highest in the strategy instruction condition: students who received supportive writing instruction apparently developed the epistemic potential of writing. Since Bereiter's skill system is about writing development in general, we might thus even infer that strategy instruction enhanced students' writing proficiency in a broader sense.

3.6 Measuring Writing Proficiency

We invariably measured students' discipline-specific writing proficiency in our research through disciplinary writing tasks. When assessing students' texts, we involved subject teachers: after all, we wanted students' texts to

improve in the eyes of the subject teacher, i.e., on subject-specific criteria. During the implementation of the intervention, however, we noticed that subject teachers found it difficult to make the criteria for good disciplinary texts explicit. Therefore, in addition to analytical ratings, we have continuously used holistic scoring methods next to analytic scoring. Holistic scoring makes the rating task more complex: it requires raters to decide on a single score whether a text exposes different proficiency levels in the various writing components (Bacha, 2001; Barkaoui, 2010). However, the advantage of this method is that a rater can include criteria that are not covered by an analytical scoring rubric.

In two of our studies using both holistic and analytic scoring of history texts (Chapters 3 and 4), we concluded that holistic and analytic scores correlated highly, implying holistic scoring to be an appropriate method for assessment of disciplinary texts, despite teachers' knowledge of text quality criteria seeming tacit.

4. RESEARCH ASSETS AND FUTURE RESEARCH

We consider the main assets of our research to be its interdisciplinarity and its subject-specificity - which might seem paradoxical. The initial aim of our research was to build a bridge between writing skill development as taught in the L1-classroom on the one hand, and writing skill development in the disciplines on the other. Gradually, however, the differences between the subjects that at first glance seemed to be close together, turned out to be of great importance. We therefore tried to make these explicit to students. In this way, we were able to contribute to research on discipline-specificity. What is specific to writing in history and philosophy? For history, this area is less unexplored than for philosophy; however, history is often contrasted with completely different disciplines, such as mathematics and science (Goldman et al., 2016; Lee & Spratley, 2010; Shanahan & Shanahan, 2008; Shanahan et al., 2011) or compiled together with other subjects to form social studies (Graham et al., 2020). This is in fact still a collection of subjects, which have their own subtleties, though certainly relevant.

At the same time, we tried to overarch, and look for the similarities between subjects, in order to make statements for disciplinary writing in general. We tried to identify the boundaries between intermediate and disciplinary literacy (Shanahan & Shanahan, 2008). Where does generic writing end, and where does discipline-specific writing begin? Because basic literacy underlies disciplinary literacy, we were able to make good use of frequently

conducted research on general writing skills (Graham & Harris, 2018), and build on these recommendations based on subject-specific reading and writing processes (Chapter 2) and subject-specific reasoning processes (e.g., Wineburg, 1991).

Research outcomes are known to give rise to new questions. First, our study calls for theoretical research into subject-specific reading and writing process models. In the think-aloud study (Chapter 3), we obtained conceptual understanding of disciplinary reading and writing processes, but we did not develop a framework. With larger-size studies, exploring processes of a significant number of students while performing a variety of disciplinary tasks, a model could be created to serve as a basis for disciplinary literacy instruction. Frameworks for subject-specific reasoning, such as the one for historical reasoning of Van Drie and Van Boxtel (2008), can serve as a foundation for reading and writing in history. However, such frameworks will need transfer to the domain of literacy. Furthermore, to our knowledge, such a framework is not available for philosophy.

There is also a need for research on the nuance differences between different subjects (cf. Goldman et al., 2016; Shanahan & Shanahan, 2008; Shanahan et al., 2011). In addition, research on the relationship between learning and writing could be further explored across disciplines. Galbraith's dual-process model (1999) might provide its groundwork. In this model, writing is not seen as a form of problem solving, as Hayes and Flower's classic model (1980) assumes, where the writer is supposed to set goals and pursue them step by step. Content discovery is largely ignored by this model; the content is already there. The dual-process model, on the other hand, assumes that writing is inherently about discovering knowledge or content. Writing involves the interplay of two processes: the knowledge-constituting process, in which implicit knowledge is activated and comes to the surface, and the knowledge-transforming process, in which the knowledge is organized, aimed at communicating the knowledge.

Another question left unanswered by our research is the extent to which students actually modified their writing strategies as a result of strategy instruction. We did offer students two different routes in the strategy instruction (for "free writers" and "pre-planners"), however, we did not investigate which writing routines were preferred by students, and whether, after instruction, they applied a different strategy route. Furthermore, from previous research (Van Steendam et al., 2022) on writing routines, we know that students might change their routines from task to task, depending on their topic

knowledge and topic interest. Further research might determine what could be the additional influence of disciplines. Relatedly, in the current project, we did not consider the extent to which the dual-route option was actually utilized or appreciated by students. This should be addressed in a subsequent iteration. Finally, our research raises the question to what extent results are sustainable.

Our research also provokes more practice-oriented questions. For example, what writing genres and writing strategies would be useful for learning in other subjects, such as geography and economics (e.g., proceeding work of Sampson, et al., 2013a, 2013b; Levin, et al., 2021)? What would strategies and instruction look like for other grades and levels? In our research, we already empowered teachers to take an active role in developing materials, but the strategy and instruction were designed by us, since we were the experts with regard to writing processes and writing instruction, which was precisely what teachers lacked. But what might be achieved if we also left the design of strategies and strategy instruction to teachers? Previous intervention studies with large teacher involvement have shown promising results (Van Drie et al., 2017; Van Drie et al., 2021).

5. PRACTICAL IMPLICATIONS AND RECOMMENDATIONS

Our research adds to the current discussion about students' declining literacy proficiency, and methods for how to convert this trend. As a response to the Dutch Education Inspectorate's (2020) advice for each school to have a "language policy", many schools have been working on the development of strategies with regard to students' literacy. Increasingly, school leaders become aware of literacy as a shared responsibility, and initiate school-wide strategies. However, the awareness that language development is not only the responsibility of the language teacher has yet to register among content teachers.

In our studies, we promote a literacy view that might go even further than the notice that literacy is a shared responsibility. We recommend not only practicing generic language skills in content lessons, but also to primarily focus on discipline-specific literacy skills instead. Each content teacher is responsible for students' literacy development in their discipline. From our research project, several recommendations have arisen, which we will explain in the next section.

5.1 Deliberate Writing Task Design

Inclusion of reading-writing tasks will be necessary to create opportunities for literacy skill development. For subject teachers to contribute to students' development of literacy, they should think carefully about their task design: reading-writing tasks should fit the learning goals teachers have in mind for their students. When tasks are appropriate for the intended content learning, they will not be "extra" but merely "alternative". A writing task becomes a learning activity in itself, rather than just a way of processing new knowledge, which had already been acquired through another learning activity. It is likely that teachers will prefer to use tasks which prompt students to include reasoning into their texts, instead of tasks which only require the outcome of a reasoning process.

Recommendation 1: Include Writing-to-Learn Tasks

The first prerequisite for developing students' literacy skills is to create opportunities to do so. Reading-writing tasks that can promote learning or comprehension of subject matter are ideally suited for content lessons: in this way, learning and developing literacy converge. A first requirement for a good writing-to-learn task is that the student's reasoning is activated by a stimulating prompt, evoking different angles toward an issue (e.g., an evaluative question). A second requirement is that students have not yet been fully instructed on the issue. Background knowledge might be required, but if an issue has already been extensively covered by the teacher in-class, there is no need for the learner to further investigate, since the brainwork has essentially already been done.

Recommendation 2: Select Sources Deliberately

Writing requires knowledge; therefore, we recommend the use of source texts. In this way, reading skills can be trained functionally: students need the information from the sources to be able to write a text. It is important that the source texts represent different perspectives on the given issue. Sources are thus advised to partly complement or contradict each other. Primary source texts are preferable to secondary texts, since primary documents are more likely to point students' attention towards sourcing (evaluating usefulness and trustworthiness of a source) and enhance the use of references in students' texts (Rouet et al., 1996). However, secondary sources might be preferred in some cases, depending on the writing task goal (cf. Van Driel et al., 2022a). An important point, is that the number of sources needs to be limited, in order to

keep the students' attention focused on the essence, and to keep the task manageable within the given time frame.

Recommendation 3: Use Short Writing Tasks

To guide student writing processes, writing must take place in-class, within the teacher's sight, which can only be accomplished with short tasks. Encompassing yet contained writing tasks provide a manageable workload for teachers, and do not impede student motivation, as students may dread reading and writing long texts. Moreover, short tasks facilitate frequent use, and regular writing promotes writing skills (Graham & Harris, 2017). However, this latter finding was not confirmed in this thesis; in the Chapter 4 study, a condition with writing tasks, but without strategy instruction, did not outperform a control condition without writing. To conclude: simply adding writing tasks into the curriculum will thus not improve students' writing nor their learning, as it needs to be accompanied by support.

5.2 Support Students' Processes

Subject teachers need efficient and effective teaching methods to support students' writing processes. Providing students with strategies is therefore recommended, as a strategy breaks a process down into smaller steps, making the "cognitive load" of the complex literacy task more manageable (Galbraith & Torrance, 2004; Torrance & Galbraith, 2006). Self-regulated strategy development (SRSD) is appropriate for all types of learning, but it has been most extensively studied in the area of writing (Harris & Graham, 2015). According to the SRSD framework, a strategy is best described, then modeled, and then practiced.

In order to describe and model a process, teachers need to know what processes are actually required. This can be challenging; teachers are not always aware of their own processes, let alone their strategies, and what other strategies might be effective. It is therefore important for teachers to become aware of the processes at play in their subjects, the variation between subjects, and possible strategies for addressing these processes. Theoretical research can provide a foundation on which to build instruction.

Recommendation 4: Provide and Model Strategies

It is recommended to provide direct, explicit instruction on required reading and writing processes (Graham & Perin, 2007). After all, it seems unfair to expect that students are able to write a text genre which they have never

written before. In addition to making the strategy explicit, it is also important to show how the strategy can be applied. Modeling (live or using a video) gives students the opportunity to learn through observation. Observational learning has proven its service for writing development in previous research (Braaksma et al., 2004).

Recommendation 5: Highlight Disciplinary Aspects

It is imperative to explicitly highlight disciplinary aspects of the process and of the product. This requires knowledge of cognitive processes in general, and awareness of subject-specific processes in particular. Professional development in this area is important: a professional learning community with colleagues from the same subject (to examine the specificity of a subject) or with colleagues from a variety of subjects (to be able to expose differences with other subjects) can be useful here. Teachers then construct knowledge by comparing and contrasting.

Recommendation 6: Support Students' Writing During Task Performance

It is recommended to support students' process during reading and writing, by providing feedback on students' reading and writing while they are in the middle of it. This can be achieved by writing in-class, within the sight of the teacher, instead of instructing students to write a text as homework. During writing, the teacher might walk through the classroom to observe and to scaffold when needed.

Recommendation 7: Use Exemplars

The process of discussing exemplars of different levels, is a type of teacher feedback. Through series of example texts, it becomes clear which dimensions of text quality are in play, how they may interact in texts, and what are the options for, a discipline-specific text. This learning activity is in fact a first step in formative assessment: a didactic approach to stimulate students' self-regulatory skills. Discussing exemplars are aimed to clarify teacher expectations of student writing, which can be seen as the first phase in the Formative Assessment Cycle (Gulikers & Baartman, 2017).

Recommendation 8: Connect Product Feedback to the Process

If product feedback is provided, this feedback can be linked to the student's process. Do you think a student's writing needs more audience focus? Explain why, and then relate it to the process steps in the writing stage, e.g.: "This

particular student needs to pay more attention to taking the reader's role after the text is written. They will need to learn to put themselves in the reader's shoes, to reread and check: did my message come across?" The writer may need the help of a peer reader, to read the text aloud, and maybe adding question marks in the margin when comprehension becomes difficult or unclear. This process may result in a reorganized second draft. If the student had a choice of strategies, the chosen strategy might also be discussed: should the student try a different route or improve the chosen route?

5.3 Support Teachers in Integrating Writing Instruction

We also pose two recommendations regarding external aids, which might help future content teachers to include writing support in their lessons.

Recommendation 9: Develop Teachers' Pedagogical Content Knowledge of Writing

Content teachers' pedagogical content knowledge with regard to (disciplinary) writing might become a learning objective in teacher education. During their teacher education, teachers learn about general and disciplinary writing processes, effective reading and writing instruction, and process support. Our studies have shown that teachers are well capable of designing sound writing-to-learn tasks based on design principles. However, the other components (development of the strategy and the instructional unit) were provided by us. In the future, teachers will need to develop their own instructional units, and to be able to their purposes. This requires deeper knowledge about processes, and experience with how students confront particular reading-writing tasks. These aspects of literacy instruction in content classrooms deserve a place in the teacher education curriculum, to properly equip future teachers to develop students' disciplinary literacy.

We recommend teacher support to focus on a number of knowledge components: (1) knowledge about writing tasks, which type of task is appropriate when; (2) knowledge about text quality, especially disciplinary aspects; (3) knowledge about (disciplinary) writing processes; and (4) procedural knowledge about how to guide writing processes, and (5) how to discuss writing products.

Recommendation 10: Include Writing Support in Textbooks

Textbook publishers might also take a role in developing students' literacy skills. For example, workbooks might need to be supplemented with writing-

to-learn tasks, which can be accompanied by process support, for example by including strategies for students and modeling videos, or by stimulating pre-writing or re-writing activities.

6. CONCLUSION

With this thesis, we have tried to answer the research question: Which instructional approach is best suited to develop students' disciplinary literacy effectively and efficiently, in upper secondary history and philosophy education? Based on four studies, we conclude that secondary school teachers who include disciplinary reading-writing tasks into their curriculum, and who provide additional support to students' writing process by means of strategy instruction, support the development of students' disciplinary literacy skills. Our studies also provide insights in teachers' interaction with design principles for writing task design, and show what we can achieve if we involve teachers in designing and implementing interventions in general.

REFERENCES

- Alger, C. L. (2007). Engaging student teachers' hearts and minds in the struggle to address (il)literacy in content area classrooms. *Journal of Adolescent & Adult Literacy*, 50(8), 620-630. <https://doi.org/10.1598/JAAL.50.8.1>
- Alkema, E. G. (2022). *Schrijven met kennis van zaken: Een didactiek voor de zaakvakken* [Knowledgeable writing. Writing instruction for social sciences]. Doctoral dissertation. University of Amsterdam.
- Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research*, 54, 577-596.
- Baaijen, V. M., & Galbraith, D. (2018). Discovery through writing: Relationships with writing processes and text quality. *Cognition and Instruction*, 36(3), 199-223. <https://doi.org/10.1080/07370008.2018.1456431>
- Bacha, N. (2001). Writing evaluation: What can analytic versus holistic essay scoring tell us? *System*, 29(3), 371-383. [https://doi.org/10.1016/S0346-251X\(01\)00025-2](https://doi.org/10.1016/S0346-251X(01)00025-2)
- Bangert-Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research*, 63(1), 69-93. <https://doi.org/10.3102/00346543063001069>.
- Barkaoui, K. (2010). Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly*, 7(1), 54-74. <https://doi.org/10.1080/15434300903464418>
- Bazerman, C. (1981). What written knowledge does: Three examples of academic discourse. *Philosophy of the Social Sciences*, 11, 361-388. <https://doi.org/10.1177/004839318101100305>
- Bereiter, C. (1980). Development in writing. In L. W. Gregg & E. R. Steinberg (Ed.), *Cognitive processes in writing* (pp. 73-93). Hillsdale, NJ: Lawrence Erlbaum.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.
- Berger, R. (2015). Now I see it, now I don't: researcher's position and reflexivity in qualitative research. *Qualitative Research*, 15(2), 219-234.
- Biancarosa, C., & Snow, C. E. (2006). *Reading next. A vision for action and research in middle and high school literacy: a report to Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.
- Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review*, 30, 703-725.
- Braaksma, M. A., Rijlaarsdam, G., Van den Bergh, H., & Van Hout-Wolters, B. H. M. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction*, 22(1), 1-36.
- Borren, M. (2012). *Handleiding filosofische leesvaardigheden* [Manual for philosophical reading]. Rijksuniversiteit Groningen.
- Brante, E., & Strømsø, H. I. (2018). Sourcing in text comprehension: A review of interventions targeting sourcing skills. *Educational Psychology Review*, 30, 773-799. <https://doi.org/10.1007/s10648-017-9421-7>
- Bråten, I., & Strømsø, H. I. (2011). Measuring strategic processing when students read multiple texts. *Metacognition and Learning*, 6(2), 111-130.
- Breetvelt, I., Van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality: When and how? *Cognition and Instruction*, 12(2), 103-123.

- Britt, M. A., & Aglinskias, C. (2002). Improving students' ability to identify and use source information. *Cognition and instruction*, 20(4), 485-522. https://doi.org/10.1207/S1532690XCI2004_2
- Britt, M. A., & Rouet, J. F. (2012). Learning with multiple documents: Component skills and their acquisition. In J. R. Kirby, & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (pp. 276-314). Cambridge University Press.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42. <https://doi.org/10.3102/0013189X018001032>
- Casado-Ledesma, L., Cuevas, I., Van den Bergh, H., Rijlaarsdam, G., Mateos, M., Granado-Peinado, M., & Martín, E. (2021). Teaching argumentative synthesis writing through deliberative dialogues: instructional practices in secondary education. *Instructional Science*, 49(4), 515-559. <https://doi.org/10.1007/s11251-021-09548-3>
- Carless, D., & Chan, K. K. H. (2017). Managing dialogic use of exemplars. *Assessment & Evaluation in Higher Education*, 42(6), 930-941. <https://doi.org/10.1080/02602938.2016.1211246>
- Carter, M. (2007). Ways of knowing, doing, and writing in the disciplines. *College Composition and Communication*, 58(3), 385-418.
- Chen, Y. C., Park, S., & Hand, B. (2016). Examining the use of talk and writing for students' development of scientific conceptual knowledge through constructing and critiquing arguments. *Cognition and Instruction*, 34(2), 100-147.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161-238). Mahwah, NJ: Lawrence Erlbaum Associates.
- Coffin, C. (2006). Learning the language of school history: the role of linguistics in mapping the writing demands of the secondary school curriculum. *Journal of Curriculum Studies*, 38(4), 413-429. <https://doi.org/10.1080/00220270500508810>
- Concepción, D. W. (2004). Reading philosophy with background knowledge and metacognition. *Teaching Philosophy*, 27(4), 351-368.
- Corcelles, M., & Castelló, M. (2017). Learning philosophical thinking through collaborative writing in secondary education. *Journal of Writing Research*, 7(1), 157-200. <https://doi.org/10.17239/jowr-2015.07.01.07>
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research Methods in Education*. Seventh Edition. London: Routledge.
- CvTE, College voor Toetsen en Examens [Dutch Institute for Assessment and Exams] (2022). *Syllabus centraal examen 2022. Geschiedenis VWO* [Syllabus central exam 2022. History Pre-University Track]. Utrecht: College voor Toetsing en Examens.
- CvTE, College voor Toetsen en Examens [Dutch Institute for Assessment and Exams] (2018). *Examen geschiedenis vwo 2018-I* [History exam 2018-I]. <https://www.examenblad.nl/examen/geschiedenis-vwo-3/2018>
- De La Paz, S. (2005). Effects of historical reasoning instruction and writing strategy mastery in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology*, 97, 139-156. <https://doi.org/10.1037/0022-0663.97.2.139>
- De La Paz, S., & Felton, M. K. (2010). Reading and writing from multiple source documents in history: Effects of strategy instruction with low to average high school writers. *Contemporary Educational Psychology*, 35, 174-192. <https://doi.org/10.1016/j.cedpsych.2010.03.001>
- De La Paz, S., Malkus, N., Monte-Sano, C., & Montanaro, E. (2011). Evaluating American history teachers' professional development: Effects on student learning. *Theory & Research in Social Education*, 39(4), 494-540. <https://doi.org/10.1080/00933104.2011.10473465>
- De La Paz, S., Felton, M., Monte-Sano, C., Croninger, R., Jackson, C., Shim Deogracias, J., & Polk Hoffman, B. (2014). Developing historical reading and writing with adolescent readers: Effects on student learning. *Theory and Research in Social Education*, 42(2), 228-274. <https://doi.org/10.1080/00933104.2014.908754>

- De La Paz, S. & Graham, S. (2002). Explicitly teaching strategies, skills and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology*, 94, 687-698. <https://doi.org/10.1037//0022-0663.94.4.687>
- De La Paz, S., Monte-Sano, C., Felton, M., Croninger, R., Jackson, C. & Worland Piantedosi, K. (2017). A historical writing apprenticeship for adolescents: Integrating disciplinary learning with cognitive strategies. *Reading Research Quarterly*, 52(1), 31-52. <https://doi.org/10.1002/rrq.147>
- De Oliveira, L. C. (2011). *Knowing and writing school history. The language of students' expository writing and teachers' expectations*. Charlotte, NC: Information Age Publishing.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, 38(3), 181-199. <http://www.jstor.org/stable/20532527>
- De Vries, G. (1984). *De ontwikkeling van wetenschap. Een inleiding in de wetenschapsfilosofie* [The development of science. An introduction to philosophy of science]. Wolters-Noordhoff.
- Dutch Education Inspectorate (2022). *De staat van het onderwijs 2022* [The State of Education 2022]. Dutch Ministry of Education, Culture and Science. <https://www.onderwijsinspectie.nl/documenten/rapporten/2022/04/13/de-staat-van-het-onderwijs-2022>
- Elbow, P. (1973). *Writing without teachers*. Oxford: Oxford University Press.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215-251. <https://doi.org/10.1037//0033-295x.87.3.215>
- Ferrari, M., Bouffard, T., & Rainville, L. (1998). What makes a good writer? Differences in good and poor writers' self-regulation of writing. *Instructional Science*, 26(6), 473-488, <https://doi.org/10.1023/A:1003202412203>
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365-387. <https://doi.org/10.2307/356600>
- Foxworth, L., & Mason, L. H. (2017). Writing-to-learn instruction that works. In R. Fidalgo & T. Olive (Series Eds.) & R. Fidalgo, K. R. Harris, & M. Braaksma (Vol. Eds.), *Studies in Writing Series: Vol. 34. Design Principles for Teaching Effective Writing* (pp. 66-86). Leiden: Brill.
- Freedman, S. W., Hull, G. A., Higgs, J. M., & Booten, K. P. (2016). Teaching writing in a digital and global age: Toward access, learning, and development for all. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 1389-1450). American Educational Research Association. https://doi.org/10.3102/987-0-935302-48-6_23
- Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science*, 21, 45-72. <https://doi.org/10.1007/BF00119655>
- Galbraith, D. (1999). Writing as a knowledge constituting process. In G. Rijlaarsdam, E. Espéret, M. Torrance, & D. Galbraith (Eds.), *Studies in Writing: Vol. 4. Knowing What to Write: Conceptual Processes in Text Production* (pp. 139-160). Amsterdam: Amsterdam University Press.
- Galbraith, D. (2009). Writing as discovery. Teaching and Learning Writing. BJEP Monograph Series II, 6. *British Journal of Educational Psychology*, 2(6), 5-26.
- Galbraith, D. (2015). Conditions for writing to learn. *Journal of Writing Research*, 7(1), 215-226. <https://doi.org/10.17239/jowr-2015.07.01.09>
- Galbraith, D., & Baaijen, V. M. (2018). The work of writing: Raiding the inarticulate. *Educational Psychologist*, 53(4), 238-257. <https://doi.org/10.1080/00461520.2018.1505515>
- Galbraith, D., & Torrance, M. (2004). Revision in the context of different drafting strategies. In L. Allal, L. Chanquoy and P. Largy (Eds.). *Revision: Cognitive and instructional processes* (pp. 63-86). Dordrecht, NL: Kluwer Academic Publishers.
- Geisler, C. (1994). *Academic literacy and the nature of expertise: Reading, writing, and knowing in academic philosophy*. Hillsdale, NJ: Lawrence Erlbaum. <https://doi.org/10.4324/9780203812174>
- Gillespie, A., Graham, S., Kiuahara, S., & Hebert, M. (2014). High school teachers use of writing to support students' learning: A national survey. *Reading and Writing*, 27(6), 1043-1072. <https://doi.org/10.1007/s11145-013-9494-8>

- Gillis, V. (2014). Disciplinary literacy: Adapt not adopt. *Journal of Adolescent & Adult Literacy*, 57(8), 514-623. <https://doi.org/10.1002/JAAL.301>
- Goldberg, A., Russell, M., Cook, A., & Russell, E. M. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment*, 2(1), 1-52. <https://ejournals.bc.edu/index.php/jtla/article/view/1661>
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M. A., Greenleaf, C., Lee, C. D., Shanahan, C., & Project READI (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 51(2), 219-246. <https://doi.org/10.1080/00461520.2016.1168741>
- Goldman, S. R. (2012). Adolescent literacy: Learning and understanding content. *The Future of Children*, 22(2), 89-116. <https://www.jstor.org/stable/23317413>
- Graham, S. (2020). The sciences of reading and writing must become more fully integrated. *Reading Research Quarterly*, 55(S1), S35-S44. <https://doi.org/10.1002/rrq.332>
- Graham, S., Capizzi, A., Harris, K. R., Hebert, M., & Morphy, P. (2014). Teaching writing to middle school students: A national survey. *Reading and Writing*, 27, 1015-1042. <https://doi.org/10.1007/s11145-013-9495-7>
- Graham, S., & Harris, K. R. (2017). Evidence-based writing practices: A meta-analysis of existing meta-analyses. In R. Fidalgo & T. Olive (Series Eds.) & R. Fidalgo, K. R. Harris, & M. Braaksma (Vol. Eds.), *Studies in Writing Series: Vol. 34. Design Principles for Teaching Effective Writing* (pp. 13-37). Leiden: Brill.
- Graham, S., & Harris, K.R. (2018). Evidence-based writing practices: A meta-analysis of existing meta-analyses. In R. Fidalgo, & T. Olive (Series Eds.) & R. Fidalgo, K.R. Harris, & M. Braaksma (Eds.), *Studies in Writing Series: Vol. 34. Design Principles for Teaching Effective Writing* (pp. 13-37). Leiden: Brill.
- Graham, S. & Hebert, M. (2010). Writing to read: Evidence for how writing can improve reading. A report from Carnegie Corporation of New York.
- Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, 81(4), 710-744.
- Graham, S., Kiuahara, S. A., & MacKay, M. (2020). The effects of writing on learning in science, social studies, and mathematics: A meta-analysis. *Review of Educational Research*, 90(2), 179-226. <https://doi.org/10.3102/0034654320914744>
- Graham, S., Liu, X., Aitken, A., Ng, C., Bartlett, B., Harris, K. R., & Holzapfel, J. (2018). Effectiveness of literacy programs balancing reading and writing instruction: A meta-analysis. *Reading Research Quarterly*, 53(3), 279-304. <https://doi.org/10.1002/rrq.194>
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445-476.
- Graham, S., Schwartz, S. S., & MacArthur, C. A. (1993). Knowledge on writing and the composing process, attitude toward writing, and self-efficacy for students with and without learning disabilities. *Journal of Learning Disabilities*, 26(4), 237-249. <https://doi.org/10.1177/002221949302600404>
- Gulikers, J., & Baartman, L. (2017). *Doelgericht professionaliseren: Formatieve toetspraktijken met effect! Wat DOET de docent in de klas?* [Targetted professional development: Formative assessment practices with effect! What the teacher DOES in the classroom]. NRO Review Report No.: 405-15-722. https://www.nro.nl/sites/nro/files/migrate/Inhoudelijke-eindrapport_NRO-PPO-405-15-722_DEF.pdf
- Hand, B., Prain, V., & Wallace, C. (2002). Influences of writing tasks on students' answers to recall and higher-level test questions. *Research in Science Education*, 32(1), 19-34.
- Handley, K., & Williams, L. (2011). From copying to learning: Using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education*, 36(1), 95-108

- Harris, K., & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation* (2nd ed.). Cambridge: Brookline Books.
- Harris, K., & Graham, S. (2009). Self-regulated strategy development in writing: Premises, evolution, and the future. *Teaching and Learning Writing*, 113–135. <https://doi.org/10.1348/978185409X422542>
- Harris, K. R., & Graham, S. (2015). Self-regulated strategy development: Theoretical bases, critical instructional elements, and future research. In R. Fidalgo & T. Olive (Series Eds.) & R. Fidalgo, K. R. Harris, & M. Braaksma (Vol. Eds.) *Studies in Writing Series: Vol 34. Design principles for teaching effective writing*. Leiden: Brill.
- Hayes, J. R., & Flower, L. S. (1981). Identifying the organisation of writing processes. In L. W. Gregg, & E. R Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. *American Psychologist*, 41(10), 1106–1113. <https://doi.org/10.1037/0003-066X.41.10.1106>
- Hebert, M., Gillespie, A., & Graham, S. (2013). Comparing effects of different writing activities on reading comprehension: A meta-analysis. *Reading and Writing*, 26, 111–138. <https://doi.org/10.1007/s11145-012-9386-3>
- Heller, R., & Greenleaf, C. L. (2007). *Literacy instruction in the content areas: Getting to the core of middle and high school improvement*. Alliance for Excellent Education.
- Ho, A., Watkins, D., & Kelly, M. (2001). The conceptual change approach to improving teaching and learning: An evaluation of a Hong Kong staff development programme. *Higher Education*, 42, 143-169. <https://doi.org/10.1023/A:1017546216800>
- Hof, M., Braaksma, M., & Rijlaarsdam, G. (2015). *Learning during writing: A methodology to unravel learning processes* [unpublished manuscript]. Amsterdam: University of Amsterdam.
- Holdinga, C. C. (2013). *Zaakvakdocenten en hun visie(s) op het ontwikkelen van de schrijfvaardigheid van havo/vwo-bovenbouwleerlingen*. Scriptie in het kader van de postmasteropleiding Academisch Meesterschap. [Content teacher views on writing skill development of upper secondary school students. Post-master's thesis]. Amsterdam: University of Amsterdam.
- Hunt, K. W. (1965). Grammatical structures written at three grade levels. NCTE Research Report No. 3.
- Kember, D. (1997). A reconceptualisation of the research into university academics' conceptions of teaching. *Learning and Instruction*, 7(3), 255–275. [https://doi.org/10.1016/S0959-4752\(96\)00028-X](https://doi.org/10.1016/S0959-4752(96)00028-X)
- Kieft, M., Rijlaarsdam, G., & Van den Bergh, H. (2008). An aptitude-treatment interaction approach to writing-to-learn. *Learning and Instruction*, 18(4), 379-390. <https://doi.org/10.1016/j.learninstruc.2007.07.004>
- Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, 11(3), 203-270. <https://doi.org/10.1023/A:1021913217147>
- Klein, P.D., & Boscolo, P. (2016). Trends in research on writing as a learning activity. *Journal of Writing Research*, 7(3), 311- 350. <https://doi.org/10.17239/jowr-2016.07.03.01>
- Koek, A. (2020). Filosofisch lezen [Philosophical reading]. In: D. Berendsen, N. Kienstra, K. Poortier, & F. Rombout (Eds.) *Filosofie op school. Handboek vakdidactiek filosofie* (pp. 225-235). Amsterdam: Boom.
- Krippendorff, K. (2013). *Content analysis. An introduction to its methodology*. Third Edition. LA: Sage Publications.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Kuhn, D., Hemberger, L., & Khait, V. (2015). Dialogic argumentation as a bridge to argumentative thinking and writing. *Journal for the Study of Education and Development*, 39(1), 25-48. <https://doi.org/10.1080/02103702.2015.1111608>
- Langer, J. A., & Applebee, A. N. (1987). *How Writing Shapes Thinking: A Study of Teaching and Learning*. NCTE Research Report No. 22. National Council of Teachers of English.

- Le Coultre, E.-A., Dooremalen, H. & Jongenelen, B. (2013). *Ik denk / Cogito. Inleiding in de filosofie. Werkboek vwo* [I think/ Cogito. Introduction to philosophy. Workbook pre-university level]. Veen Media.
- Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy*. Final report from Carnegie Corporation of New York's Council on advancing adolescent literacy. Carnegie Corporation of New York.
- Lenski, S., & Johns, J. L. (1997). Patterns of reading-to-write. *Reading Research and Instruction*, 37(1), 15-38. <https://doi.org/10.1080/19388079709558252>
- Levin, D. M., De La Paz, S., Lee, Y., & Escola, E. N. (2021). Use of cognitive apprenticeship models of instruction to support middle school students' construction and critique of written scientific explanations and arguments. *Learning Disabilities: A Multidisciplinary Journal*, 26(1), 58-72. <https://doi.org/10.18666/LDMJ-2021-V26-I1-10380>
- Mac Mahon, B. (2014). Making the invisible visible: Disciplinary literacy in secondary school classrooms. *Irish Educational Studies*, 33(1), 21-36. <https://doi.org/10.1080/03323315.2013.867243>
- Marsman, P. (2010). *Vakdossier filosofie* [Subject file philosophy]. Enschede: Stichting leerplanontwikkeling (SLO).
- McDermott, M., & Hand, B. (2013). The impact of embedding multiple modes of representation within writing tasks on high school students' chemistry understanding. *Instructional Science*, 41(1), 217-246.
- Martínez, I., Mateos, M. D. M., Martín, E., & Rijlaarsdam, G. (2015). Learning history by composing synthesis texts: Effects of an instructional programme on learning, reading and writing processes, and text quality. *Journal of Writing Research*, 7(2), 275-302. <https://doi.org/10.17239/jowr-2015-07-02-03>
- Mateos, M., Martín, E., Villalón, R., & Luna, M. (2008). Reading and writing to learn in secondary education: online processing activity and written products in summarizing and synthesizing tasks. *Reading and Writing*, 21(7), 675-697. <https://doi.org/10.1007/s11145-9086-6>
- McCarthy Young, K. M., & Leinhardt, G. (1998). Writing from primary documents: A way of knowing in history. *Written Communication*, 15(1), 25-68. <https://doi.org/10.1177/0741088398015001002>
- McKinley, W. (1992). The role of reading and writing while composing from sources. *Reading Research Quarterly*, 27(3), 227-248.
- McKenney, S., & Reeves, T. C. (2019). *Conducting Educational Design Research*. Second Edition. Oxon/New York: Routledge.
- Merchie, E., & Van Keer, H. (2014). Using on-line and off-line measures to explore fifth and sixth graders' text-learning strategies and schematizing skills. *Learning and Individual Differences*, 32, 193-203. <https://doi.org/10.1016/j.lindif.2014.03.012>
- Mierwald, M., Lehmann, T., & Brauch, N. (2022). Writing about the past: The impact of different authentic instructional material on students' argument writing in history. *European Journal of Psychology of Education*, 37, 163-184. <https://doi.org/10.1007/s10212-021-00541-5>
- Miller, D. M., Scott, C. E., & McTigue, E. M. (2018). Writing in the secondary-level disciplines: a systematic review of context, cognition, and content. *Educational Psychology Review*, 30, 83-120. <https://doi.org/10.1007/s10648-016-9393-z>
- Ministry of Education, Culture, and Science (2009). *Referentiekader Taal en Rekenen* [Reference Framework for Language and Math]. <https://www.rijksoverheid.nl/onderwerpen/taal-en-rekenen/referentiekader-taal-en-rekenen>
- Moje, E. B. (2008). Foregrounding the disciplines in secondary teaching and learning: A call for change. *Journal of Adolescent & Adult Literacy*, 52, 97-107. <https://doi.org/10.1598/JAAL.52.2.1>
- Moje, E. B., Ciechanowski, K. M., Kramer, K., Ellis, L., Carrillo, R., & Collazo, T. (2004). Working toward third space in content area literacy: An examination of everyday funds of knowledge and discourse. *Reading Research Quarterly*, 39(1), 38-70. <https://doi.org/10.1598/RRQ.39.1.4>

- Monte-Sano, C. (2008). Qualities of historical writing instruction: A comparative case study of two teachers' practices. *American Educational Research Journal*, 45(4), 1045-1079. <http://www.jstor.org/stable/27667162>
- Monte-Sano, C. (2010). Disciplinary literacy in history: An exploration of the historical nature of adolescents' writing. *Journal of Learning Sciences*, 19(4), 539-568. <https://doi.org/10.1080/10508406.2010.481014>
- Monte-Sano, C. (2011). Beyond reading comprehension and summary: Learning to read and write in history by focusing on evidence, perspective and interpretation. *Curriculum Inquiry*, 41(2), 212-249. <https://doi.org/10.1111/j.1467-873X.2011.00547.x>
- Monte-Sano, C., & Allen, A. (2018). Historical argument writing: the role of interpretive work, argument type, and classroom instruction. *Reading and Writing*, 32(6), 1383-1410. <https://doi.org/10.1007/s11145-018-9891-0>
- Monte-Sano, C., & De La Paz, S. (2012). Using writing tasks to elicit adolescents' historical reasoning. *Journal of Literacy Research*, 44(3) 273-299. <https://doi.org/10.1177/1086296X12450445>
- Monte-Sano, C., De La Paz, S., & Felton, M. (2014). Implementing a disciplinary-literacy curriculum for US history: learning from expert middle school teachers in diverse classrooms. *Journal of Curriculum Studies*, 46(4), 540-575. <https://doi.org/10.1080/00220272.2014.904444>
- Mottart, A., Vanbrabant, P., & Van de Ven, P.-H. (2009). Schrijven bij diverse schoolvakken. Een verkenning [Writing in different subjects. An exploration]. *Levende Talen Tijdschrift*, 10(4), 14-21.
- National Governors Association & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors. Retrieved from <https://learning.ccsso.org/wp-content/uploads/2022/11/ADA-Compliant-ELA-Standards.pdf>
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, 51(2), 102-116.
- Newell, G. E. & Winograd, P. (1995). Writing about and learning from history texts: The effects of task and academic ability. *Research in the Teaching of English*, 19(2), 133-163. <https://www.jstor.org/stable/40171246>
- Nieveen, N. (1999). Prototyping to reach product quality. In J. van den Akker, R.M. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds), *Design approaches and tools in education and training* (pp. 125-136). Boston: Kluwer Academic.
- Nokes, J. D. (2010). Observing literacy practices in history classrooms. *Theory & Research in Social Education*, 38(4), 515-544. <https://doi.org/10.1080/00933104.2010.10473438>
- Nokes, J. D., & De La Paz, S. (2018). Writing and argumentation in history education. In Metzger, S. A., & McArthur Harris, L. (Eds.) *The Wiley international handbook of history teaching and learning*, 551-578. New York: Wiley-Blackwell.
- Nokes, J. D., Dole, J. A., Hacker, D. K. (2007). Teaching high school students to use heuristics while reading historical texts. *Journal of Educational Psychology*, 99(3), 492-504. <https://doi.org/10.1037/0022-0663.99.3.492>
- O'Brien, D. G., Stewart, R. A., & Moje, E. B. (1995). Why content literacy is difficult to infuse into the secondary school: Complexities of curriculum, pedagogy, and school culture. *Reading Research Quarterly*, 30(3), 442-463. <https://doi.org/10.2307/747625>
- O'Donnell, C. L. (2008) Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, 78(1), 33-84. <https://doi.org/10.3102/0034654307313793>
- OECD (2018). *PISA 2018 Results: Combined executive summaries. Volume I, II & III*. https://www.oecd.org/pisa/Combined_Executive_Summaries_PISA_2018.pdf
- Olinghouse, N. G., Graham, S., & Gillespie, A. (2015). The relationship of discourse and topic knowledge to fifth graders' writing performance. *Journal of Educational Psychology*, 107(2), 391-406. <https://doi.org/10.1037/a0037549>
- Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education* 27, 309-23.

- Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly*, 19(2), 139-158.
<https://doi.org/10.1080/10573560308222>
- Pajares, F., & Valiante, G. (1997). Influence of self-efficacy on elementary students' writing. *The Journal of Educational Research*, 90(6), 353-360.
<https://doi.org/10.1080/00220671.1997.10544593>
- Polanyi, M. (1973). *Personal knowledge*. London: Routledge/Kegan Paul.
- Pressley, M. (2002). Metacognition and self-regulated comprehension. In A. E. Farstrup, & S. J. Samuels (Eds.), *What research has to say about reading instruction* (third edition), pp. 291-309.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.
- Pressley, M., & Harris, K. R. (2006). Cognitive strategies instruction: From basic research to classroom instruction. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 265-286). Lawrence Erlbaum Associates Publishers.
- Ragland, R.G. (2007). Changing secondary teachers' views of teaching American history. *The History Teacher*, 40(2), 219-246.
- Reisman, A. (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction*, 30(1), 86-112.
<https://doi.org/10.1080/07370008.2011.634081>
- Reynolds, G. A., & Perin, D. (2009). A comparison of text structure and self-regulated writing strategies for composing from sources by middle school students. *Reading Psychology*, 30, 265-300. <https://doi.org/10.1080/02702710802411547>
- Rijlaarsdam G., Janssen, T., Rietdijk, S., & Van Weijen, D. (2017). Reporting design principles for effective instruction of writing: Interventions as constructs. In R. Fidalgo & T. Olive (Series Eds.) & R. Fidalgo, K. R. Harris, & M. Braaksma (Vol. Eds.), *Studies in Writing Series: Vol. 34. Design Principles for Teaching Effective Writing* (pp. 280-313). Leiden: Brill.
- Rijlaarsdam, G., & Van den Bergh, H. (1996). An agenda for research into an interactive compensatory model of writing: Many questions, some answers. In C. M. Levy, & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 107-126). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rijlaarsdam, G., & Van den Bergh, H. (2006). Writing process theory: A functional dynamic approach. In C. A. Arthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research*. New York: The Guilford Press.
- Rogiers, A., Merchie, E., & Van Keer, H. (2019). What they say is what they do. Comparing task-specific self-reports, think-aloud protocols, and study traces for measuring secondary school students' text-learning strategies. *European Journal of Psychology of Education*, 35, 315-332. <https://doi.org/10.1007/s10212-019-00429-5>
- Rijlaarsdam, G. & Janssen, T. (1996). How do we evaluate the literature curriculum? About a social frame of reference. In E. Marum (Ed.), *Children and Books in the Modern World: Contemporary Perspectives on Literacy* (pp. 75-98). London: The Falmer Press.
- Rouet, J. F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text Relevance and Learning from Text* (pp. 19-52). Greenwich, CT: Information Age Publishing.
- Rouet, J. F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology*, 88(3), 478-493.
<https://doi.org/10.1037/0022-0663.88.3.478>
- Sampson, V., Enderle, P., Grooms, J., & Witte, S. (2013a). Writing to learn by learning to write during the school science laboratory: Helping middle and high school students develop argumentative writing skills as they learn core ideas. *Science Education*, 97(5), 643-670.
<https://doi.org/10.1002/sce.21069>
- Sampson, V., Enderle, P., & Grooms, J. (2013b). Argumentation in science education. *The Science Teacher*, 80(5), 30.

- Samuelowicz, K., & Bain, J.D. (1992). Conceptions of teaching held by academic teachers. *Higher Education*, 24, 93-111. <https://www.jstor.org/stable/3447619>
- Sanders-Reio, J., Alexander, P. A., Reio Jr., T. G., & Newman, I. (2014). Do students' beliefs about writing relate to their writing self-efficacy, apprehension, and performance? *Learning and Instruction*, 33, 1-11. <https://doi.org/10.1016/j.learninstruc.2014.02.001>
- Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming. *Advances in Applied Psycholinguistics: Volume 2, Reading, Writing, and Language Learning*, 2, 142.
- Schellings, G., Aarnoutse, C., & Van Leeuwe, J. (2006). Third-grader's think-aloud protocols: Types of reading activities in reading an expository text. *Learning and Instruction*, 16(6), 549-568. <https://doi.org/10.1016/j.learninstruc.2006.10.004>
- Schlepppegrell, M., & De Oliveira, L. C. (2006). An integrated language and content approach for history teachers. *Journal of English for Academic Purposes*, 5, 254-268. <http://www.doi.org/10.1016/j.jeap.2006.08.003>
- Schrijvers, M., Janssen, T., Fialho, O., & Rijlaarsdam, G. (2019). Toward the validation of a literature classroom intervention to foster adolescents' insight into human nature: an iterative design process. *L1-Educational Studies in Language and Literature*, 19(4), 1-47. <https://doi.org/10.17239/L1ESLL-2019.19.04.02>
- Schumacher, G. M., & Gradwohl Nash, J. (1991). Conceptualizing and measuring knowledge change due to writing. *Research in the Teaching of English*, 25(1), 67-96.
- Sendur, K. A. (2021). *Promoting written historical reasoning among undergraduate L2 students*. Doctoral Dissertation. University of Amsterdam.
- Shanahan, T., & Shanahan, C. (2017). Disciplinary literacy: Just the facts. *Educational Leadership*, 74(5), 18-22.
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40-59. <https://doi.org/10.17763/haer.78.1.v62444321p602101>
- Shanahan, C., Shanahan, T., & Mischia, C. (2011). Analysis of expert readers in three disciplines: History, mathematics, and chemistry. *Journal of Literacy Research*, 43(4), 393-429. <https://doi.org/10.1177/1086296X11424071>
- Siebert, D., & Draper, R. (2008). Why content-area literacy messages do not speak to mathematics teachers: A critical content analysis. *Literacy Research and Instruction*, 47(4), 229-245. <https://doi.org/10.1080/19388070802300314>
- Spivey, N. N. & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, 24(1), 7-26.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgements of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743-762. doi.org/10.1037/a0027627
- Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds). *Handbook of writing research* (pp. 67-80). New York: The Guilford Press.
- Torrance, M., Thomas, G. V., & Robinson, E. J. (1994). The writing strategies of graduate research students in the social sciences. *Higher Education*, 27, 379-392. <https://doi.org/10.1007/BF03179901>
- Tozzi, M. (2012). Une approche par compétences en philosophie? [A competency based approach in philosophy?]. *Rue Descartes*, 1, 22-51. <https://doi.org/10.3917/rdes.073.0022>
- Van Boxtel, C., & Van Drie, J. (2018). Historical reasoning: Conceptualizations and educational applications. In S. A. Metzger, & L. McArthur Harris (Eds.), *The Wiley international handbook of history teaching and learning*, (pp. 149-176). New York: Wiley-Blackwell.
- Van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, 54(5-6), 360-372. <https://doi.org/10.1080/0163853X.2017.1306677>

- Van den Bergh, H., & Rijlaarsdam, G. (1999). The dynamics of idea generation during writing: An online study. In M. Torrance, L. Waes, & D. Galbraith (Eds.), *Writing and Cognition* (pp. 125-150). Amsterdam, Netherlands: Elsevier.
- Vandermeulen, N. (2020). *Synthesis writing in upper-secondary education: From a baseline of texts and processes to process-oriented feedback*. Dissertation KU Leuven, Belgium.
- Vandermeulen, N., De Maeyer, S., Van Steendam, E., Lesterhuis, M., Van den Bergh, H. & Rijlaarsdam, G. C. W. (2020). Mapping synthesis writing in various levels of Dutch upper-secondary education: A national baseline study on text quality, writing process and students' perspectives on writing. *Pedagogische Studiën*, 97(2), p. 187-236.
- Vandevelde, S., Van Keer, H., Schellings, G., & Van Hout-Wolters, B. (2015). Using think-aloud protocol analysis to gain in-depth insights into upper primary school children's self-regulated learning. *Learning and Individual Differences*, 43, 11-30. <https://doi.org/10.1016/j.lindif.2015.08.027>
- Van Drie, J., Groenendijk, T., Braaksma, M., & Janssen, T. (2016). *Genrespecifiek schrijven in de mens- en maatschappijvakken* [Genre-specific writing in social studies]. Amsterdam: Landelijk Expertisecentrum Mens- en Maatschappijvakken.
- Van Drie, J., & Van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review*, 20(2), 87-110. <https://doi.org/10.1007/s10648-007-9056-1>
- Van Drie, J., Van Boxtel, C., & Braaksma, M. (2014). Writing to engage students in historical reasoning. In G. Rijlaarsdam (Series Ed.) & P. D. Klein, P. Boscolo, L. C. Kirkpatrick, & C. Gelati (Vol. Eds.), *Studies in Writing: Vol. 28, Writing as a learning activity* (pp. 94-119). Leiden: Brill.
- Van Drie, J., Van Boxtel, C., & Van der Linden, J. L. (2006). Historical reasoning in a computer-supported collaborative learning environment. In A. M. O'Donnell, C. E. Hmelo, & G. Erkens, (Eds.), *Collaborative Learning, Reasoning and Technology* (pp. 265-296). Mahwah, NJ: Lawrence Erlbaum Associates.
- Van Drie, J., Braaksma, M., & Van Boxtel, C. (2015). Writing in history: Effects of writing instruction on historical reasoning and text quality. *Journal of Writing Research*, 7(1), 123-156. <https://doi.org/10.17239/jowr-2015.07.01.06>
- Van Drie, J., Janssen, T., & Groenendijk, T. (2017). Improving writing in social studies through professional development: Effects on teachers' beliefs, classroom practice and students' writing. *European Journal of Applied Linguistics*. <https://doi.org/10.1515/eujal-2017-0009>
- Van Drie, J., Janssen, T., & Groenendijk, T. (2018). Effects of writing instruction on adolescents' knowledge of writing and text quality in history. *L1-Educational Studies in Language and Literature*, 18(2), 1-28. <https://doi.org/10.17239/L1ESLL-2018.18.03.08>
- Van Drie, J., & Stoel, G. (2020). 'Maar wat vind je hier nu goed aan?' Samen redeneringen van leerlingen analyseren als vorm van professionalisering [What do think is good about it? Collaborative assessment of students' texts as a way of professionalization]. *VELON Tijdschrift voor Lerarenopleiders*, 41(2), 87-100.
- Van Drie, J., Van Driel, J., & Van Weijen, D. (2021). Developing students' writing in history: Effects of a teacher-designed domain-specific writing instruction. *Journal of Writing Research*, 13(2), 201-229. <https://doi.org/10.17239/jowr-2021.13.02.01>
- Van Driel, J., Van Drie, J., & Van Boxtel, C. (2022a). Struggling with historical significance: Reasoning, reading, and writing processes. *European Journal of Applied Linguistics*, 10(1), 185-211. <https://doi.org/10.1515/eujal-2021-0004>
- Van Driel, J., Van Drie, J., & Van Boxtel, C. (2022b). Writing about historical significance: The effects of a reading-to-write instruction. *International Journal of Educational Research*, 112, 101924. <https://doi.org/10.1016/j.ijer.2022.101924>
- Van Ockenburg, L., Van Weijen, D., & Rijlaarsdam, G. (2021a). Learning how to synthesize: The design and evaluation of a reading-writing learning unit for high-school students. *L1-Educational Studies in Language and Literature*, 21(1), 1-33. <https://doi.org/10.17239/L1ESLL-2021.21.01.06>

- Van Ockenburg, L., Van Weijen, D., & Rijlaarsdam, G. (2021b). Choosing how to plan informative synthesis texts: Effects of strategy-based interventions on overall text quality. *Reading and Writing*, 36, 997-1023. <https://doi.org/10.1007/s11145-021-10226-6>
- Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method. A practical guide to modelling cognitive processes*. London, England: Academic Press London.
- Van Steendam, E., Vandermeulen, N., De Maeyer, S., Lesterhuis, M., Van den Bergh, H., & Rijlaarsdam, G. (2022). How students perform synthesis tasks: an empirical study into dynamic process configurations. *Journal of Educational Psychology*, 114(8), 1773-1800. <https://doi.org/10.1037/edu0000755>
- Van Veen, K., Zwart, R., & Meirink, J. (2012). What makes teacher professional development effective? A literature review. In: Kooy, M., & Van Veen, K. (Eds.). *Teacher learning that matters: International perspectives*. Vol. 62, Chapter, 3, pp. 3-22. New York: Routledge.
- Verhavert, S., Bouwer, R., Donche, V., & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education Principles Policy and Practice*, 26(2), 1-22 <https://doi.org/10.1080/0969594X.2019.1602027>
- Veenman, M. V. (2011). Learning to self-monitor and self-regulate. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197-218). New York: Routledge.
- Velema, F., & Groza, T. (2020). Het schrijven van een filosofisch essay [Writing a philosophical essay]. In: D. Berendsen, N. Kienstra, K. Poortier, & F. Rombout (Eds.) *Filosofie op school. Handboek vakdidactiek filosofie* (pp. 215-223). Amsterdam: Boom.
- Villalón, R., Mateos, M., & Cuevas, I. (2015). High school boys' and girls' writing conceptions and writing self-efficacy beliefs: what is their role in writing performance? *Educational Psychology*, 35(6), 653-674. <https://doi.org/10.1080/01443410.2013.836157>
- Villanueva, M. G., & Hand, B. (2011) Science for all: Engaging students with special needs in and about science. *Learning Disabilities Research & Practice*, 26(4), 233-240.
- Voss, J. F., & Wiley, J. (1997). Developing understanding while writing essays in history. *International Journal of Educational Research*, 27, 255-265. [https://doi.org/10.1016/S0883-0355\(97\)89733-9](https://doi.org/10.1016/S0883-0355(97)89733-9)
- Westbroek, H., Janssen, F., Mathijssen, I., & Doyle, W. (2020). Teachers as researchers and the issue of practicality. *European Journal of Teacher Education*, 1-17. <https://doi.org/10.1080/02619768.2020.1803268>
- White, M. J., & Bruning, R. (2005). Implicit writing beliefs and their relation to writing quality. *Contemporary Educational Psychology*, 30(2), 166-189. <https://doi.org/10.1016/j.cedpsych.2004.07.002>
- Wiley, J., Steffens, B., Britt, M. A., & Griffin, T. D. (2014). Writing to learn from multiple-source inquiry activities in history. In G. Rijlaarsdam (Series Ed.) & P.D. Klein, P. Boscolo, L. C. Kirkpatrick, & C. Gelati (Vol. Eds), *Studies in Writing: Vol. 28, Writing as a learning activity* (pp. 120-148). Leiden: Brill.
- Wiley, J., & Voss, J. F. (1996). The effects of 'playing historian' on learning in history. *Applied Cognitive Psychology*, 10(7), 63-72.
- Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, 91(2), 301. <https://doi.org/10.1037/0022-0663.91.2.301>
- Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 73-87. <https://doi.org/10.1037/0022-0663.83.1.73>
- Wineburg, S. S. (1994). The cognitive representation of historical texts. In J. Castellan, D. B. Pisoni, & G. Potts (Eds.), *Teaching and Learning in History* (pp. 85-135). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wissinger, D. R., Ciullo, S. P. & Shiring, E. J. (2018). Historical literacy instruction for all learners: Evidence from a design experiment. *Reading & Writing Quarterly*, 34(6), 568-586. <https://doi.org/10.1080/10573569.2018.1510795>

- Yin, R. K., (1994). *Case study research design and methods: Applied social research and methods series*. Thousand Oaks, CA: Sage Publications Inc.
- Yore, L. D., Bisanz, G. L., & Hand, B. M. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education*, 25, 689-725.

AUTHOR INDEX

- Aarnoutse, C., 207
Afflerbach, P., 33, 184, 206
Aglinskas, C., 84, 200
Aitken, A., 202
Alexander, P. A., 206, 207, 209
Alger, C. L., 101, 199
Alkema, E. G., 25, 199, 259, 260, 266
Allen, A., 18, 19, 74, 205
Applebee, A. N., 24, 32, 115, 199, 203
- Baaijen, V. M., 17, 62, 141, 163, 199, 201
Baartman, L., 195, 202
Bacha, N., 190, 199
Bain, J.D., 134, 207
Bangert-Drowns, R. L., 97, 199
Barkaoui, K., 190, 199
Barnard, Y. F., 209
Bartlett, B., 202
Bazerman, C., 25, 33, 199
Bereiter, C., 34, 62, 68, 103, 189, 199, 207
Berger, R., 64, 199
Biancarosa, C., 100, 199
Bisanz, G. L., 100, 210
Bisra, K., 37, 199
Booten, K. P., 201
Borren, M., 141, 199
Boscolo, P., 33, 34, 203, 208, 209
Bouffard, T., 201
Bouwer, R., 209
Braaksma, M. A., 142, 195, 199, 201, 202, 203, 206, 208, 259, 265
Brante, E., 34, 37, 73, 171, 199, 239, 249
Bråten, I., 33, 165, 199
Brauch, N., 204
- Breetvelt, I., 37, 199
Britt, M. A., 74, 76, 77, 84, 107, 138, 141, 200, 202, 206, 209
Brown, J. S., 78, 200
Brown, W., 202
Bruning, R., 68, 69, 104, 209
- Capizzi, A., 202
Carless, D., 163, 200
Carrillo, R., 204
Carter, M., 18, 33, 66, 200
Casado-Ledesma, L., 115, 200
Castelló, M., 34, 37, 140, 144, 171, 200, 239, 249
Chan, K. K. H., 163, 200
Chen, Y. C., 134, 200
Chi, M. T. H., 33, 37, 200
Ciechanowski, K. M., 204
Ciullo, S. P., 209
Coffin, C., 18, 25, 102, 200
Cohen, L., 81, 200
Collazo, T., 204
Collins, A., 200, 219
Concepción, D. W., 34, 133, 140, 200
Cook, A., 202
Corcelles, M., 34, 37, 140, 144, 171, 200, 239, 249
Cribb, G., 202
Croninger, R., 200, 201
Cuevas, I., 200, 209, 260
- De La Paz, S., 4, 11, 25, 27, 34, 68, 73, 76, 78, 84, 103, 107, 116, 126, 134, 142,

- 169, 173, 178, 187, 200, 201, 204, 205,
237, 240, 247, 251
- De Maeyer, S., 208, 209
- De Oliveira, L. C., 25, 32, 71, 72, 101,
103, 201, 207
- De Vries, G., 135, 167, 183, 201
- Desimone, L. M., 26, 134, 138, 144, 165,
166, 201
- Dole, J. A., 205
- Donche, V., 209
- Dooremalen, H., 204
- Doyle, W., 209
- Draper, R., 100, 207
- Duguid, P., 200
- Elbow, P., 141, 201
- Ellis, L., 204
- Enderle, P., 206
- Ericsson, K. A., 64, 201
- Escola, E. N., 204
- Felton, M. K., 34, 68, 73, 76, 78, 103, 134,
142, 169, 200, 201, 205
- Ferrari, M., 33, 34, 63, 88, 160, 165, 201
- Fialho, O., 207
- Flower, L. S., 25, 33, 78, 171, 191, 201,
203, 239, 249
- Foxworth, L., 25, 201, 237, 247
- Freedman, S. W., 32, 35, 201
- Galbraith, D., 4, 17, 25, 62, 67, 141, 163,
178, 191, 194, 199, 201, 207, 208
- Geisler, C., 68, 201
- George, M. A., 202
- Gillespie, A., 24, 32, 201, 203, 205
- Gillis, V., 25, 74, 102, 182, 202, 237, 247
- Goldberg, A., 97, 202
- Goldman, S. R., 22, 23, 100, 103, 132,
133, 139, 164, 166, 190, 191, 202, 237,
247
- Gradwohl Nash, J., 207
- Graham, S., 17, 19, 21, 22, 24, 25, 32, 67,
69, 74, 78, 103, 104, 107, 110, 132, 138,
142, 187, 190, 191, 194, 201, 202, 203,
205, 206, 207
- Granado-Peinado, M., 200
- Greenleaf, C. L., 133, 202, 203
- Griffin, T. D., 209
- Groenendijk, T., 208, 266
- Grooms, J., 206
- Groza, T., 137, 209
- Gulikers, J., 195, 202
- Hacker, D. K., 205
- Hand, B. M., 35, 100, 134, 169, 200, 202,
204, 209, 210
- Handley, K., 163, 164, 202
- Harris, K. R., 19, 25, 33, 62, 74, 78, 110,
138, 142, 191, 194, 201, 202, 203, 205,
206, 207
- Hayes, J. R., 25, 33, 78, 171, 191, 201,
203, 239, 249
- Hebert, M., 17, 21, 24, 25, 201, 202, 203
- Helder, A., 33, 60, 62, 141, 207
- Heller, R., 133, 203
- Hemberger, L., 203
- Higgs, J. M., 201
- Ho, A., 134, 203
- Hof, M., 37, 171, 203, 239, 249
- Holdinga, C. C., 2, 3, 25, 31, 65, 99, 131,
203, 257, 259, 260, 261
- Holzapfel, J., 202
- Hull, G. A., 201
- Hunt, K. W., 117, 203
- Jackson, C., 200, 201
- Janssen, F., 209
- Janssen, T. M., 31, 65, 136, 206, 207, 208,
257, 259, 260, 265
- Johns, J. L., 34, 204
- Jongenelen, B., 204
- Kaiser, J., 207

- Kelly, M., 203
 Kember, D., 134, 162, 203
 Khait, V., 203
 Kieft, M., 75, 77, 138, 203, 265
 King, J. R., 33, 207, 219
 Kirppendorf, K., 203
 Kiuhara, S. A., 201, 202
 Klein, P. D., 24, 32, 33, 34, 67, 125, 203, 208, 209
 Koek, A., 133, 203, 266
 Kramer, K., 204
 Kuhn, D., 37, 203
- Langer, J. A., 115, 203
 Le Coultre, E.-A., 204, 220
 Lee, C. D., 22, 190, 202, 204
 Lee, Y., 204
 Lehmann, T., 204
 Leinhardt, G., 35, 68, 84, 204
 Lenski, S., 34, 204
 Lesterhuis, M., 208, 209
 Levin, D. M., 192, 204
 Liu, Q., 199
 Liu, X., 202
 Luna, M., 204, 260
- Mac Mahon, B., 101, 204
 MacArthur, C. A., 202, 207
 MacKay, M., 202
 Malkus, N., 200
 Manion, L., 200
 Marsman, P., 138, 204
 Martín, E., 200, 204, 260
 Martínez, I., 80, 144, 204, 260
 Mason, L. H., 25, 201, 237, 247
 Mason, R. A., 206
 Mateos, M. D. M., 34, 37, 64, 200, 204, 209
 Mathijsen, I., 209
 McCarthy Young, K. M., 35, 68, 84, 204
 McDermott, M., 134, 204
 McGinley, W., 34, 204
- McKenney, S., 27, 69, 135, 204
 McTigue, E. M., 204
 Meirink, J., 209
 Merchie, E., 37, 204, 206
 Merry, S., 205
 Mierwald, M., 204
 Miller, D. M., 100, 204
 Mischia, C., 207
 Moje, E. B., 22, 25, 35, 66, 101, 102, 132, 169, 180, 204, 205, 237, 247
 Möller, J., 207
 Montanaro, E., 200
 Monte-Sano, C., 18, 19, 25, 68, 73, 74, 76, 100, 103, 107, 125, 133, 134, 200, 201, 205
 Morphy, P., 202
 Morrison, K., 200
 Mottart, A., 24, 32, 66, 205
- Nelson, T. O., 37, 205
 Nesbit, J. C., 199
 Newell, G. E., 18, 32, 67, 73, 74, 106, 205
 Newman, I., 207
 Ng, C., 202
 Nieveen, N., 69, 205
 Nokes, J. D., 103, 205
- Olinghouse, N. G., 187, 205
 Orsmond, P., 144, 205
- Pajares, F., 68, 105, 206
 Park, S., 200
 Perfetti, C. A., 206
 Perin, D., 32, 76, 78, 84, 104, 194, 202, 206
 Polanyi, M., 144, 206
 Polk Hoffman, B., 200
 Prain, V., 202
 Pressley, M., 33, 62, 184, 206
- Ragland, R.G., 206
 Rainville, L., 201

- Re, A. M., 266
Reeves, T. C., 27, 69, 135, 204
Reiling, K., 205
Reio Jr., T. G., 207
Reisman, A., 103, 206
Reynolds, G. A., 76, 78, 84, 206
Rietdijk, S., 206
Rijlaarsdam, G. C. W., 4, 31, 42, 63, 65,
99, 131, 136, 186, 199, 200, 201, 203,
204, 206, 207, 208, 209, 245, 255, 257,
259, 260
Robinson, E. J., 207
Rogiers, A., 37, 206
Rouet, J. F., 74, 76, 77, 107, 138, 141,
193, 200, 206
Russell, E. M., 202
Russell, M., 202

Salimi, F., 199
Sampson, V., 192, 206
Samuelowicz, K., 207
Sandberg, J. A. C., 209
Sanders-Reio, J., 68, 126, 207
Scardamalia, M., 34, 62, 68, 199, 207
Schellings, G., 37, 207, 208
Schleppegrell, M., 103, 207
Schrijvers, M., 66, 207, 259, 266
Schumacher, G. M., 128, 189, 207
Schwartz, S. S., 202
Scott, C. E., 204
Sendur, K. A., 178, 207
Shanahan, C., 22, 23, 24, 25, 132, 182,
190, 191, 202, 207
Shanahan, T., 22, 23, 24, 132, 182, 190,
191, 207
Shim Deogracias, J., 200
Shiring, E. J., 209
Siebert, D., 100, 207
Simon, H. A., 64, 201
Snow, C. E., 100, 199
Spivey, N. N., 33, 207
Spratley, A., 22, 190, 204

Steffens, B., 209
Stewart, R. A., 205
Stoel, G., 166, 208
Strømsø, H. I., 33, 34, 37, 73, 165, 171,
199, 239, 249
Südkamp, A., 35, 160, 207

Thomas, G. V., 207
Torrance, M., 77, 194, 201, 207, 208
Tozzi, M., 34, 207

Valiante, G., 68, 206
Van Boxtel, C., 37, 182, 191, 207, 208
Van de Ven, P.-H., 205
Van den Bergh, H., 42, 63, 199, 200,
203, 206, 208, 209
Van den Broek, P., 33, 60, 62, 141, 207
Van der Linden, J. L., 208
Van Drie, J. P., 18, 26, 27, 32, 35, 37, 65,
73, 84, 96, 99, 103, 107, 131, 166, 182,
191, 192, 193, 207, 208, 259, 260
Van Driel, J., 26, 103, 193, 208
Van Hout-Wolters, B. H. M., 199, 208
Van Keer, H., 37, 204, 206, 208
Van Leeuwe, J., 207
Van Ockenburg, L., 66, 75, 208, 209
Van Someren, M. W., 184, 209
Van Steendam, E., 191, 208, 209, 265
Van Veen, K., 137, 138, 144, 166, 209
Van Weijen, D., 206, 208, 209
Vanbrabant, P., 205
Vandermeulen, N., 85, 116, 120, 159,
208, 209, 266
Vandevelde, S., 37, 208
Veenman, M. V., 37, 209
Velema, F., 137, 209
Verhavert, S., 41, 209
Villalón, R., 104, 126, 204, 209
Villanueva, M. G., 134, 209
Voss, J. F., 18, 25, 73, 74, 84, 103, 107,
209

Wallace, C., 202
Watkins, D., 203
Westbroek, H., 26, 72, 96, 138, 209
White, M. J., 68, 69, 104, 209
Wiley, J., 18, 25, 35, 68, 73, 74, 84, 102,
103, 107, 205, 207, 209
Williams, L., 163, 164, 202
Wineburg, S. S., 25, 34, 73, 77, 182, 191,
209
Winne, P. H., 199, 206
Winograd, P., 18, 32, 67, 73, 74, 106, 205
Wissinger, D. R., 187, 209
Witte, S., 206, 265
Worland Piantedosi, K., 201

Yin, R. K., 151, 210
Yore, L. D., 100, 210

Zwart, R., 209

APPENDICES

APPENDIX A	History and Philosophy Assignment	219
APPENDIX B	Scoring Rubric	221
APPENDIX C	Questionnaires Learner Variables	223
APPENDIX D	Model Comparisons (Chapter 3)	225
APPENDIX E	Interview Guides	227
APPENDIX F	Factorial Analysis Self-Efficacy Questionnaire	231
APPENDIX G	Model Comparisons (Chapter 4)	233

APPENDIX A: HISTORY AND PHILOSOPHY ASSIGNMENT

History task

In his Ecclesiastical History of the English People, completed around 731, the Venerable Bede (a monk) writes about the war that broke out in 654 between the rulers of two Anglo-Saxon kingdoms, Oswiu of Bernicia and Penda of Mercia:

Text 1

About this time, King Oswiu was exposed to the ferocious and unstoppable attacks of Penda, the king of Mercia, who had killed Oswiu's brother. Ultimately, Oswiu was forced to promise him a huge portion of the royal treasury in exchange for peace. The condition was that Penda would return home and stop destroying Oswiu's kingdom. But the heathen king would not accept this offer, for he was determined to exterminate the entire nation, from high to low. Oswiu appealed to God's grace and help, seeing that nothing else could save him and his people from this barbaric and ruthless enemy. (...) Thus he prepared for battle with his small army. It is said that the Gentile army was thirty times larger. (...) The battle began and the Gentiles were put to flight and destroyed. The thirty leaders who fought alongside Penda almost all lost their lives.

Based on: Bede (Author), Collins, R. & McClure, J. (Eds.) (2008). *The Ecclesiastical History of the English People* (pp. 149-150). Oxford: Oxford University Press

Suppose you are writing a paper on the history of Anglo-Saxon England. You want to use this source in two chapters: one on the course of the power struggle between Bernicia and Mercia, and one on the world view of a monk in an Anglo-Saxon monastery.

Assignment

Referring to the source text, identify for which chapter this source text contains less useful information and for which chapter it contains more useful information.

This assignment was derived from the Dutch College voor Toetsen en Examens (2018).

Philosophy task

In evolutionary biology, human morality can be explained from the theory of evolution. According to evolutionary biologist Richard Dawkins, our genes are 'selfish', and we are 'programmed' to pass on our genes rather than look out for our own survival as individuals. The following is an excerpt from *Philosophy Magazine*, from an interview with British science journalist Matt Ridley (a supporter of Dawkins):

Text 1

"When I take care of my children, I serve the self-interest of my genes. But that does not mean that I am pursuing my individual self-interest, rather the opposite. It only costs me. If I were purely selfish, I wouldn't even start doing so. That explanation goes further than you might think. Take for example economics. Classical economics does not provide any explanation for the fact that people leave money to their children. According to economists, this is irrational because it does not serve self-interest. The selfish gene theory offers an explanation. When you leave money, the genes are selfish, not the individual."

Monfils, L. (2008). *Philosophy Magazine*, 5.

Assignment

Do you think people act solely in their own interest? Give an example that supports the statement that people act only out of self-interest and give an example that opposes this statement. Then provide a substantiated answer to the question of whether people act solely from self-interest.

This assignment was based on Le Coultre et al. (2013).

APPENDIX B: SCORING RUBRIC

Rubric for Assessment of Content Quality

		Score				
		1	2	3	4	5
Understanding of the issue	The text barely displays understanding of the main issue, or understanding is lacking completely		The text displays a simple and basic understanding of the main issue		The text displays a thorough understanding of the main issue.	
Multiperspectivity	The text is one-sided; the main issue is discussed, but not from different perspectives.		Opposing views are presented in the text, but not very elaboratively		This understanding exceeds the specific information in the documents; in the text, own knowledge and information from the source materials are intertwined.	
Argumentation	The arguments in the text are weak and/or barely elaborated on		The different views are supported by strong and/or weaker arguments, and the arguments are only moderately elaborated on		Opposing views are presented in the text	
Use of sources	Source information is presented as 'own text' (e.g. it is presented as facts, and/or references are not provided)		Source information is made explicit in the text by referencing (e.g. 'According to Eduard Douwes Dekker ...')		The different views are supported by strong arguments, which are elaborated on with accurate facts, examples and details	
					Source information is integrated in the text; authors of sources are not only mentioned, but also discussed (i.e., the author's position as a reporter)	

Rubric for Assessment of Structure

		Score				
		1	2	3	4	5
Global text structure	Overall, the text has no clear structure, which makes the text difficult to understand			Overall, the text has a reasonably clear and logical structure. The text is divided into paragraphs, but those are not always logical or are not presented in the right order.		Overall, the text has a clear and logical structure. The text has a strong introduction and conclusion, and is divided into clear paragraphs.
Local text structure	The text is incoherent. The text presents paragraphs or sentences in an unclear manner, because structural elements (signal words, linking words) are lacking. Because of this, the text seems to be a series of sentences without coherence.			Paragraphs are reasonably unified. Linkages between paragraphs or between sentences are mostly implicit, e.g. because linking words or signal words are lacking.		The paragraphs are unified and coherent, both internally and from paragraph to paragraph. Transitions are clear, i.e. through the use of connectors.

APPENDIX C: QUESTIONNAIRES LEARNER VARIABLES

Writing Beliefs Questionnaire

1. Good writers include a lot of quotes from authorities in their writing.
2. The most important goal of writing is to convey information clearly to the readers.
3. Good writing involves making as few revisions as possible.
4. Writing should focus on the information in books and articles.
5. The key to successful writing is accurately reporting what authorities think.
6. The most important reason to write is to report what authorities think about a subject.
7. Writing requires constantly revising to improve what has been written.
8. Writing is a process involving a lot of emotion.
9. It is important to develop a distinctive writing style.
10. Good writers deviate from the information they need for a topic. (re-code)
11. Good writing involves frequent revising.
12. Writing is often an intense experience.
13. Writing helps me to understand better what I'm thinking about.
14. I always feel that one more revision will make my text even better.
15. Writing helps me see the complexity of ideas.
16. My thoughts and ideas become more clear to me as I write and rewrite.
17. Writers' points of view should become clear from their writing.
18. Writing is often an emotional experience.
19. Writers should immerse themselves in their writing.
20. It is important to objectively represent your sources.
21. Writing is correctly representing what the sources say.
22. Writing objectively is very important to me.
23. I write to clarify what others think.
24. I write to get a grip on a subject.
25. Writing absorbs me.
26. Writing sometimes takes the utmost out of me.

Self-Efficacy Questionnaire for Source-Based Writing in History

1. I can derive from the assignment what the main historical issue is.
2. I can purposefully read a source.
3. I can analyze sources for reliability and representativeness.
4. I can organize source information.
5. I can determine whether sources contradict each other or whether they agree.
6. I can identify different perspectives in the source information.
7. I can explain different viewpoints by when and where each text was written.
8. I can identify information that is relevant for the issue.
9. I can select information that is relevant for my text.
10. I can write a text that corresponds with the source information.
11. I can make connections between information from different sources.
12. I can synthesize information from different sources into one sentence or paragraph.
13. I can explain source information in my own words.
14. I can write a text that is comprehensible for someone who has not read the source materials.
15. I can describe an historical event as an introduction for my text.
16. I can use historical concepts in my writing.
17. I can write arguments supported by evidence.
18. I can explain and illustrate source information in my own text.
19. I can write a text with a clear main idea.
20. I can write a text where all paragraphs are useful.
21. I can write a text without redundant sentences.
22. I can use conjunctions to link sentences and paragraphs.
23. I can write a well-structured text.
24. I can write a text with coherent paragraphs.
25. I can vary my sentence structures and choice of words.
26. I can write a text without errors and typos.
27. I can identify and correct language errors.
28. I can divide my time between reading and writing well.
29. I can monitor my process while performing a reading-writing task.
30. During the process, I sometimes decide to go back to an earlier stage in the process.

APPENDIX D: MODEL COMPARISONS (CHAPTER 3)

Holistic quality

Model	χ^2	df	Models	Comparison		
				χ^2	df	p
0 Intercept + 2 variance components		3				
1 Plus: Time	1017.796	5	0 vs 1		2	.627
2 Plus: Condition	1017.780	6	1 vs 2	0.016	1	.901
3 Plus: Interaction Time*Condition	1013.718	8	2 vs 3	4.062	2	.133

Mean holistic quality (and standard errors), estimated under model 3:

Condition	T1	T2	T3
Group 1 (E-C)	84.1 (3.6)	81.8 (3.9)	79.4 (3.8)
Group 2 (C-E)	82.5 (3.8)	79.0 (4.3)	87.3 (4.5)

Content

Model	χ^2	df	Models	Comparison		
				χ^2	df	p
0 Intercept + 2 variance components		3				
1 Plus: Time	282.673	5	0 vs 1		2	.110
2 Plus: Condition	282.672	6	1 vs 2	0.001	1	.977
3 Plus: Interaction Time*Condition	276.644	8	2 vs 3	6.028	2	.050

Mean content quality (and standard errors), estimated under model 3:

	T1	T2	T3
Group 1 (E-C)	2.60 (0.17)	2.36 (0.18)	2.26 (0.18)
Group 2 (C-E)	2.50 (0.18)	2.13 (0.20)	2.69 (0.21)

Structure

Model	χ^2	df	Models	Comparison		
				χ^2	df	p
0 Intercept + 2 variance components		3				
1 Plus: Time	323.145	5	0 vs 1		2	.047
2 Plus: Condition	322.214	6	1 vs 2	0.931	1	.336
3 Plus: Interaction Time*Condition	320.308	8	2 vs 3	1.906	2	.387

Mean text structure quality (and standard errors), estimated under model 3:

Condition	T1	T2	T3
Group 1 (E-C)	2.74 (0.20)	2.29 (0.22)	2.44 (0.21)
Group 2 (C-E)	2.35 (0.21)	2.01 (0.24)	2.48 (0.25)

Text length

Model	χ^2	df	Models	Comparison		
				χ^2	df	p
0 Intercept + 2 variance components		3				
1 Plus: Time	1466.586	5	0 vs 1		2	.019
2 Plus: Condition	1466.571	6	1 vs 2	0.001	1	.903
3 Plus: Interaction Time*Condition	1464.652	8	2 vs 3	6.028	2	.384

Mean text length (and standard errors), estimated under model 3:

Condition	T1	T2	T3
Group 1 (E-C)	234 (22.8)	170 (24.9)	175 (24.2)
Group 2 (C-E)	207 (24.3)	186 (27.4)	189 (28.7)

APPENDIX E: INTERVIEW GUIDES

Explorative Interviews with Teachers (Prestudy, Chapter 3)

- A. Analysis of the writing task
1. What is the reason you chose to bring this specific task to this interview? In what way is this task relevant for your subject area?;
 2. Who developed this task?;
 3. Could this task also be a part of an assessment?;
 4. What are the characteristics of the source texts? (primary/secondary, genre, length, difficulty);
 5. What are the characteristics of the text students wrote? (length, genre, audience);
 6. What is the learning goal of the task?;
 7. To what extent have students achieved this learning goal?;
 8. Were you satisfied with students' results?
- B. Analysis of students' texts
1. What is it that makes the weak text so weak?;
 2. Although you judge the weak text as weak, is there anything good in it?;
 3. What is it that makes the good text so good?;
 4. What could be improved in the good text?;
 5. To what extent is this task discipline-specific? What is typical history in this task?
- C. Support and feedback
1. What did you and the students do prior to task performance? (instruction, support) |;
 2. What did you and the students do during the task performance? (collaboration with peers, help from the teacher or from tools, time spent, questions asked);
 3. What did you and the students do after task performance? (grading or not, oral or written feedback from teacher or peers, assessment rubric available or not);
 4. In retrospect: what would you do differently prior, during or after task performance?;
 5. What might help the student who wrote the good text improve? (prior, during, after performance);
 6. What might help the student who wrote the weak text improve? (prior, during, after performance)
- D. Cognitive processes
1. What cognitive processes are involved in the task? Please write them on sticky notes;
 2. Can you describe how a student should perform the task, subsequently, from beginning to end? Please, paste the sticky notes in the right order or way to represent the process as a whole, as a flow chart;
 3. Could the job be done any other way, different from how you have described the process until now?;
 4. What is the main thing you would like to see your students do differently in their process?;
 5. How could you take care of that?
- E. Writing beliefs
1. How important is it for your students to have good writing skills?

Reflective Interviews with Teachers (Chapter 3)

- A. Descriptive
 - a. What did you do in the regular lessons? Please, provide a description of the learning activities during these lessons;
 - b. How did you conduct the intervention lessons? Did you skip or add elements? Why?
- B. Evaluative
 - a. To what extent do you think the lesson materials are practical? What improvements do you see?;
 - b. To what extent do you think the lesson materials are useful for learning? What improvements do you see?;
 - c. Which elements would you reuse?;
 - d. How did students respond to the lessons? Do you think the learning goal of writing better historical texts was achieved?

Reflective Interviews with Students (Chapter 3)

- A. Program differentiation
 - 1. Were the history lessons any different than you were used to? What was different? Was that positive or negative?
- B. Motivation
 - 2. How motivated were you during the intervention lessons? Why?;
 - 3. How motivated were your classmates?
- C. Evaluative
 - 4. How did you experience performing evaluative tasks?;
 - 5. What did you think of the intervention lessons?;
 - 6. What did you learn from the intervention lessons?;
 - 7. What would you have wanted to learn in addition?;
 - 8. What did you think of the strategy?;
 - 9. Would you keep using the strategy in future assignments?
 - 10. Which step in the strategy was most useful for you?;
 - 11. Did you use the cheat sheet while making the assignments?;
 - 12. What did you think of the video?;
 - 13. What did you think of the modeling peer in the video?;
 - 14. What did you think of the model texts?;
 - 15. What would you tell next years' students in 11th grade who are starting the intervention lessons?;
 - 16. What improvements for the intervention lessons would you suggest?

Reflective Interviews with Teachers (Chapter 4)

1. How did you implement strategy lesson 1? Did you skip or add lesson elements? Why?*
2. How did you implement strategy lesson 2? Did you skip or add lesson elements? Why?*
3. How did the students perform task 1? Why?
4. In what way was task 1 discussed in class? Why?
5. How did the students perform task 2? Why?
6. In what way was task 2 discussed in class? Why?
7. To what extent do you think the lessons* and tasks were useful? What improvements do you see?
8. Would you re-use the writing tasks?
9. Would you re-use the strategy lessons? Which elements?*
10. To what extent do you think the strategy-instruction was beneficial for students? Could it be left out?*
11. When re-using the writing tasks, would you add strategy-instruction? Why do you think this would be worthy?***
12. How did students react to the lessons* and the tasks?
13. What do you think students have learned?
14. How did you perceive developing the writing tasks?

Reflective Interviews with Students (Chapter 4)

1. Were the lessons in the previous class period different? What was different?
2. Was this a negative or a positive difference?
3. How were the writing tasks performed? Did you need help? With what? Was help provided?
4. What did you think of the writing tasks?
5. Did you perceive the writing tasks as useful for learning?
6. Did you feel efficacious to perform the tasks?***
7. How were the writing tasks discussed in class?
8. Was that discussion useful for you?
9. What did you think of the strategy lessons?*
10. What did you think of the strategy itself?*
11. Would you re-use the strategy?*
12. Which step in the strategy was most worthy to you?*
13. What did you think of the video?*
14. What did you think of the example texts?*
15. What is the most important thing you learned from the lessons?
16. What would you have wanted to learn in addition?
17. Did you perform all the tasks? How motivated were you and your classmates to perform the tasks? Why?
18. How was the class motivation during the strategy lessons?*
19. What would you say to future students who are starting the lessons* and tasks?
20. What would you advise your teacher when he/she is about to re-use the lessons* and tasks?

* Question only for Condition WT+S

** Question only for Condition WT

APPENDIX F. FACTORIAL ANALYSIS
SELF-EFFICACY QUESTIONNAIRE

Summary of Exploratory Factor Analysis Results for the Self-Efficacy Questionnaire

Statement	Rotated Factor Loadings (N = 242)					
	1	2	3	4	5	6
1	.22	-.02	-.10	.12	.52	-.16
2	.52	.06	.13	.12	.08	.18
3	.34	.00	-.08	.01	.23	-.03
4	.24	.13	.04	-.07	.39	.14
5	.58	.06	-.14	-.08	.04	.03
6	.11	.00	-.02	-.02	.47	.18
7	.05	-.04	-.05	.11	.70	-.04
8	.75	.03	.10	.16	.07	-.02
9	.80	.04	-.02	.16	-.11	-.04
10	.54	.21	-.03	-.12	.17	.10
11	.49	-.05	-.12	-.08	.25	-.00
12	.25	.17	-.08	-.03	.19	.27
13	.38	-.14	-.21	-.02	.07	.23
14	.13	.18	-.15	-.14	.37	.16
15	-.13	-.01	-.04	.09	.79	.07
16	.17	-.04	-.04	.09	.43	.11
17	.23	-.07	-.22	.03	.20	.29
18	.15	.05	-.04	-.04	.38	.32
19	.12	-.07	-.27	-.08	.12	.46
20	-.04	.03	-.09	.20	-.00	.67
21	-.08	.31	-.07	.28	.07	.29
22	-.03	-.02	-.68	-.07	.02	.11
23	.13	.03	-.76	.11	-.19	.11
24	-.04	.04	-.82	.05	.10	-.06
25	-.07	.23	-.55	.06	.15	-.07
26	.08	.88	-.05	-.01	-.06	-.09
27	-.01	.85	-.01	.01	-.05	.02
28	.02	.07	.05	.49	.26	.06
29	.25	.01	-.09	.63	-.02	.12
30	.08	.12	-.30	.44	.09	-.02
Eigenvalues	10.95	2.35	1.69	1.32	1.17	1.05
% of variance	36.51	7.82	5.64	4.39	3.89	3.51
α	.86	.87	.83	.75	.82	.68

Note: Factor loadings over .40 are marked. Statements are elaborated in Appendix C.

The items that clustered in the same factor suggested that factor 1 represented *sourcing*, factor 2 represented *language*, factor 3 *structure*, factor 4 *monitoring*, factor 5 *historical reasoning*, and factor 6 *coherence*. Correlation analyses revealed that factors 1, 3, 5 and 6 were related. Factors 2 (language) and 4 (monitoring) appeared to be more or less independent constructs.

APPENDIX G: MODEL COMPARISONS (CHAPTER 4)

Model Comparisons for Writing Proficiency Indicators

Model	df	Holistic score		Text length		Situational understanding		Multi-perspectivity		Argumentation		Source use		Structure	
		χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>
0 Intercept + 2 Un-structured components	4	4250.471	< .001	5982.417	< .001	1318.079	< .001	1357.689	< .001	1262.091	< .001	1383.766	< .001	1381.415	< .001
1 + Time	5	4243.112	.007	5973.483	.003	1310.002	.005	1352.376	.021	1253.756	.004	1342.898	< .001	1367.023	< .001
2 + Condition	7	4167.100	.001	5903.422	.001	1244.861	.001	1282.373	.006	1190.793	.001	1284.373	< .001	1299.792	< .001
3 + Time*Condition	9	4140.144	< .001	5864.882	< .001	1229.578	< .001	1261.992	< .001	1179.706	.004	1255.911	< .001	1267.515	< .001

Model Comparisons for Epistemic Experience

Model		df	χ^2	<i>p</i>
0	Intercept + 2 Unstructured components	4	1240.415	.947
	+ Time	5	1240.415	.998
2	+ Condition	7	1231.841	.884
3	+ Time*Condition	9	1225.290	.038

Model Comparisons for Course Content Knowledge Indicators

Model		df	Number of new T-units		Mean length of new T-units	
			χ^2	<i>p</i>	χ^2	<i>p</i>
0	Intercept + 2 Unstructured components	3	1193.112	< .001	1253.346	< .001
1	+ Condition	5	1184.594	.014	1245.712	.022

Model Comparisons for Holistic Score With Covariates

Covariates		χ^2	df	<i>p</i>
Transmissional beliefs				
Model 4	+ Cov	3953.433	10	.644
Model 5	+ Time*Cov	3948.182	12	.072
Model 6	+ Condition*Cov	3948.023	13	.691
Model 7	+ Time*Condition*Cov	3947.512	15	.774
Cognitive engagement beliefs				
Model 4	+ Cov	3946.463	10	.007
Model 5	+ Time*Cov	3945.929	12	.077
Model 6	+ Condition*Cov	3945.403	13	.047
Model 7	+ Time*Condition*Cov	3944.312	15	.580
Emotional engagement beliefs				
Model 4	+ Cov	3953.366	10	.060
Model 5	+ Time*Cov	3946.383	12	.030
Model 6	+ Condition*Cov	3945.870	13	.474
Model 7	+ Time*Condition*Cov	3942.921	15	.229
Self-efficacy total score				
Model 4	+ Cov	3771.236	10	.048
Model 5	+ Time*Cov	3769.215	12	.221
Model 6	+ Condition*Cov	3768.090	13	.724
Model 7	+ Time*Condition*Cov	3767.918	15	.917

Model Comparisons for Epistemic Experience With Covariates

Covariates	χ^2	df	<i>p</i>
Transmissional beliefs			
Model 4 + Cov	1163.026	10	.471
Model 5 + Time*Cov	1159.859	12	.201
Model 6 + Condition*Cov	1159.648	13	.646
Model 7 + Time*Condition*Cov	1156.596	15	.218
Cognitive engagement beliefs			
Model 4 + Cov	1151.659	10	.001
Model 5 + Time*Cov	1150.406	12	.530
Model 6 + Condition*Cov	1147.947	13	.117
Model 7 + Time*Condition*Cov	1142.983	15	.085
Emotional engagement beliefs			
Model 4 + Cov	1159.527	10	.044
Model 5 + Time*Cov	1159.193	12	.842
Model 6 + Condition*Cov	1158.674	13	.472
Model 7 + Time*Condition*Cov	1151.094	15	.022
Self-efficacy total score			
Model 4 + Cov	1112.883	10	.160
Model 5 + Time*Cov	1108.800	12	.120
Model 6 + Condition*Cov	1108.518	13	.596
Model 7 + Time*Condition*Cov	1100.608	15	.020

SUMMARY

The aim of this dissertation was to contribute to the development of secondary school students' disciplinary literacy, with a focus on writing proficiency. Students' declining literacy levels, and at the same time, society's elevated threshold, have increased the urgency to respond. In this study, we explored how teachers could address disciplinary literacy in their classrooms without major disruptions to their regular teaching practices.

1. CHAPTER 1: INTRODUCTION

Within content areas, writing is generally used as a tool for learning content or assessing students' knowledge and reasoning. Little emphasis is placed on learning-to-write. This might be obvious: teaching students how to write a good text is generally seen as the responsibility of the language teacher. Yet, embedding literacy activities in content-classes is important, given the differences across disciplines in how knowledge is created, what strategies are used for inquiry, and which genres, which concepts, and what language are involved (Goldman, et al., 2016). It is the responsibility of the content teacher to address this disciplinary literacy.

Writing instruction has been validated as effective for writing-to-learn in content classrooms (i.e., enhancing writing quality and knowledge gain) (De La Paz, 2005; Foxworth & Mason, 2017). However, subject teachers do not always have the knowledge to provide writing instruction or support like language teachers have (Gillis, 2014; Moje, 2008). The goal of this dissertation was to design, implement and test lesson materials for content teachers, to facilitate them in the development and support of secondary school students' disciplinary writing. Since the school subjects history and philosophy both have high demands regarding literacy, we decided to focus on these two disciplines. Our main research question was:

Which instructional approach is suited to develop students' disciplinary literacy effectively and efficiently in upper secondary history and philosophy education?

We investigated this research question through design research: after an exploration of the context and the literature, we formulated design principles

for writing tasks and strategy instruction. The task principles guided teachers in the development of sound writing-to-learn tasks to integrate in their lessons. The strategy instruction principles formed the basis for an instructional unit, which was tailored to each subject (history and philosophy) and tested in ecologically valid contexts in several studies.

2. OVERVIEW OF THE STUDIES

We started with a prestudy (integrated in Chapter 3). The aim was to obtain a clear picture of history and philosophy teachers' writing practice, and their needs for improvement of teaching disciplinary writing. We conducted interviews with teachers of history ($n = 10$) and philosophy ($n = 11$) from different schools. We asked them what type(s) of writing tasks they commonly used in their subject at the upper secondary level. Teachers were asked to bring a representative example of a writing task for their subject, and we used this example task as a "hook". We asked the teachers what demands they made on the product of the task, how they supported the students' writing process, and what they thought the students' process should resemble.

The prestudy showed that teachers had difficulty with the verbalization of writing processes. Both philosophy and history teachers expressed that they found it challenging to describe what students should actually do to successfully complete their assignment. The majority of history teachers interviewed disregarded aspects of the writing process; they focused on reading and evaluating sources rather than, for example, on structuring or revising their previously written text. Philosophy teachers expressed more awareness of the role of writing processes in their assignments, implying that they recognized writing as a means of learning philosophy.

2.1 Chapter 2

The subsequent study was also a background study, yet focusing on students. We studied students' processes while they were performing writing tasks for the two concerning subjects in a think-aloud study. Fifteen high-achieving students (11th grade) who took both subjects were selected, since we aimed to explore which cognitive processes novice-experts would show when they were reading and writing in the two disciplines. For each subject, we selected a writing task based on sources, which was ought to be representative of the subjects in question, based on the prestudy results. Students performed the two tasks in random order while thinking-aloud. We analyzed their patterns,

to see if and to what extent these varied per discipline, and whether the cognitive activities could explain text and thought quality.

To identify students' cognitive processes, we employed a data-driven approach to find patterns. For the segmenting and coding of students' think-aloud protocols, we also made use of existing frameworks for reading and writing in general (e.g., Hayes & Flower, 1981), and reading and writing in the disciplines (e.g., Brante & Strømsø, 2018; Corcelles & Castelló, 2015; Hof et al., 2015). Three main categories: reading, writing, and metacognition, were subdivided into 11 process activities. All process activities were analyzed on frequency, and absolute and relative duration.

Subsequently, students' protocols were assessed on two measures of quality: (1) quality of the eventual text, and (2) quality of students' thought process. Afterwards, quantitative analyses were conducted: we researched correlations between the quality of students' texts, the quality of students' thought process, and their process activities.

Results showed that the relations between quality and activities differed per task, or, as we suggested, per discipline. In the history assignment, text quality mostly depended on planning variables. Students who spent more time on planning and planned more frequently, produced higher quality texts. By contrast, in the philosophy assignment, most of the variables affecting quality were writing variables. Students who wrote relatively high-quality texts showed characteristics of expert writers: they produced longer texts, spent more time writing, and reviewed more often.

Furthermore, the relation between text quality and thought process quality differed per subject. The history assignment mainly provoked the outcome of students' reasoning process, while the philosophy assignment invited students to write down their entire reasoning as well; the philosophy task thus stimulated thinking and writing to codevelop.

The results of this contextual study had implications for the next steps in our research: we considered the character of the philosophy assignment, which stimulated students' thought process, to be a suited task type for use in other content classrooms as well. As opposed to the history assignment, which only required an outcome, the philosophy assignment could be characterized as a writing-to-learn task; such tasks are well suited for use in content classrooms, since they fit the focus on knowledge development.

2.2 Chapter 3

In a next study, design principles and an instructional design to enhance students' historical writing were developed. The instructional unit was evaluated on validity, practicality, and effectiveness. Insights from our prestudy, think-aloud study (Chapter 2), and a literature search led us to the formulation of two design principles, which formed the foundation of the instructional units for both history and philosophy. The two design principles were:

If we want students to develop a profound understanding of content through writing, then:

Design principle #1: students should write short evaluative texts, based on multiple primary sources that represent multiple perspectives.

Design principle #2: students should be provided with discipline-specific, dual-route, reading-writing strategy instruction which is easily applicable for teachers.

Strategy instruction was included to support the construction of coherent knowledge; making the writing process manageable is likely to promote knowledge construction, and having a dual route to accommodate different writing process preferences prevents the writing process from becoming a burden.

To put design principle 2 into practice, a strategy was designed, based on the insights obtained in our think-aloud study, and previous interventions studies using writing strategy instruction in the content classroom (e.g., De La Paz et al., 2017). This resulted in the seven-step Read-Think-Write Strategy (RTW strategy), which was accustomed to historical evaluative questions, in the form of "to what extent"-questions: questions that were assumed to be writing-to-learn questions. The strategy offered two routes to accommodate different writing process preferences: "free writing", for students who prefer to write from flow and will revise thoroughly, and "pre-planning", for students who prefer to start writing from a text scheme.

In a trial study, the instructional unit was implemented and evaluated by two history teachers in a switching replications design with three measurement occasions. As measurements, three different writing tasks were designed in cooperation with the participating teachers. This procedure enhanced feasibility and encouraged teachers' engagement with the materials.

Students' texts were assessed by a jury of history teachers on three criteria: holistic quality, content quality, and quality of structure. Results showed indications of effectiveness on content quality. However, results were not fully convincing, possibly because the second principle did not stand out sufficiently. The participating teachers valued the writing tasks as described in principle 1. They doubted however some aspects of the strategy instruction as mentioned in principle 2. The main recommendation arising from this trial was to increase teachers' background knowledge of the design principles and to explain the underpinnings behind these principles more prominently. In a redesign, a teacher development session was thus included.

2.3 Chapter 4

In a third study, we tested the effectiveness of the optimized instructional design in a quasi-experimental study with three research conditions, to obtain insights into effects of two factors: (1) implementation of writing tasks, and (2) implementation of writing process instruction.

In a first experimental condition (7 groups, 119 students, 6 teachers), participating teachers designed writing tasks, which were in line with design principle 1. Next to that, they implemented an optimized version of the designed writing strategy instruction, using the RTW strategy. In a second experimental condition (3 groups, 63 students, 2 teachers), teachers also designed writing tasks, but the strategy instruction was omitted. A third condition was a non-writing control condition (4 groups, 86 students, 3 teachers).

In a quasi-experimental pretest-posttest design, with writing tasks as pretest and posttest, we aimed to measure effects on students' writing proficiency. These writing tasks were evaluative tasks on a topic which was not discussed during the history lessons. After the strategy instruction – containing direct instruction, modeling through a video showing modeling peers, and discussions of exemplars – students practiced writing through performance of two writing tasks, which were designed by their teacher and therefore fully tailored to the content of their history lessons. These writing practice tasks thus differed from group to group.

A jury team of history teachers assessed students' pre- and posttest texts on five aspects: situational understanding, multiperspectivity, source use, argumentation, and structure. Furthermore, they assessed texts holistically, by means of a text scale. These scores were indicators for students' writing proficiency in history.

Furthermore, we explored the effects of the writing tasks and strategy instruction on students' epistemic experience: to what extent students experienced writing tasks as learning tasks. Finally, we explored whether writing tasks were a proper alternative to other, more common, learning activities for learning history. Therefore, we measured students' course content knowledge in a pre-post recall test. The topic of this recall test differed from group to group, as each group followed their regular year plan.

Results showed effects on students' writing proficiency: students who had received writing strategy instruction outperformed other students: they wrote better texts. This applied to students with all self-efficacy levels and belief types. Significant effects of the additional strategy instruction were visible in all aspects of text quality, holistically and on discipline-specific criteria. Furthermore, students who received strategy instruction more often indicated they experienced the epistemic function of writing, compared to the control condition.

Recall test results showed that students who received strategy instruction performed equally well as students in the other conditions regarding course content knowledge gain: time had been devoted to writing strategy instruction, and was not wasted, so was feared by content teachers. These results imply that writing seems a promising alternative activity for history learning, however, only when attention is paid to the writing process, this results in more advanced writing proficiency.

2.4 Chapter 5

To compare different disciplines, we investigated the effects of a similar instructional design for philosophy in a fourth study. This study focused on the interaction of three philosophy teachers with writing task design and the instructional unit, which they implemented into their philosophy lessons in 10th-grade. The instructional design of the history unit was adapted to philosophical writing, and open to context modifications. Implementation was accompanied by teacher guidance activities, which were aimed to prompt teachers' contemplation of writing instruction and writing quality.

This study can be characterized as a design study, combined with a multiple-case study, to profoundly explore how philosophy teachers interacted with the designed materials and what results it yielded for students' disciplinary writing proficiency.

Three philosophy teachers were instructed to design their own writing tasks for use in their 10th grade classes and implemented the designed

instructional unit. We monitored and evaluated teachers' implementation, to explore how resilient the design would be in open, ecologically valid contexts and conducted evaluative interviews with students and teachers to research their adjustments to the design.

Furthermore, we explored student progress as perceived by teachers. Students' writing proficiency was measured with three writing tasks, two of which were designed by the group's teachers, and thus differing from group to group. Teachers then selected a benchmark text for each task: the "average" text for that task. This activity elicited reflection on quality criteria in philosophical writing.

To frame teachers' perceptions of progress, the writing tasks and students' texts written in response to these tasks were assessed by independent jury teams, with benchmark texts as references. In reflective interviews teachers explained and reflected on their students' scores.

Results indicated that philosophy teachers were able to design qualified writing-to-learn tasks, when instructed on theoretically grounded design principles. The accompanying instructional unit turned out to be feasible and instructive. With regard to students' progress, we carefully concluded that the instructional unit was a valuable contribution to students' philosophical writing proficiency. Students showed more independent philosophical thinking in their text after the intervention, whilst tasks became more complex. The teacher guidance program elicited reflection on the value of writing tasks and instruction for subject goals, and on perceptions of student progress.

In general, the results of Chapter 5 provoke discussion about what writing tasks and writing instruction can provide for philosophy education and what a high-quality philosophical text precisely entails.

3. DISCUSSION

3.1 Conceptual Issues

Three lines of development did become apparent through the four studies of this dissertation, namely, with regard to (1) disciplinary writing; (2) supporting writing; and (3) innovation of teachers' writing practice.

With this dissertation, we advocate the integration of writing tasks that can promote learning in order to work on students' disciplinary literacy. Learning is generally enhanced when students are prompted to explore different perspectives on an issue, by means of multiple sources to analyze and interpret, and to provide their own perspectives on the issue based on the sources, in an

evaluative text. Results of our studies showed that teachers valued the evaluative writing tasks as learning activities, history teachers in particular. For philosophy teachers, the recommended short length of the tasks tore into teachers' conceptions of the learning effects of writing assignments. Nevertheless, teachers recognized that short tasks can be a way of utilizing writing for learning, namely, to master the procedure of philosophical writing.

Our studies have shown that disciplines differ: writing processes as well as texts have discipline-specific elements, which should be addressed in content classes. However, our prestudy revealed that teachers have limited knowledge about writing processes, which is an obstacle for providing writing instruction and support. Therefore, we developed feasible materials, such as a video with modeling peers, to relieve teachers of the complex, knowledge-requiring components of writing instruction, which was valued. The main idea of process instruction was considered a worthy addition to the content classroom.

Implementation of the designed unit also presented challenges. Along the way, we experienced that teacher guidance was needed to successfully integrate and implement the designed materials. A remaining challenge for teachers was the discussion of exemplars. To address subject specificity in a writing task and text, a teacher will need to be aware of what a good historical or philosophical text actually entails, which was not always straightforward. However, teacher guidance focusing on the evocation of reflection seemed to be helpful.

Regarding innovation of teachers' practice, we recommend to design materials open to context modifications, as this flexibility enhances practicality and feasibility. In this way, we avoid the creation of barriers that might prevent teachers from integrating writing instruction after all.

3.2 Methodological Issues

Next, we discussed methodological issues. We concluded that the combination of different research methods illuminated different perspectives on disciplinary writing. Quantitative measurements brought us the recognition that writing instruction is effective for development of students' writing proficiency and content knowledge. Qualitative methods gave us insight into teachers' experiences in implementing the innovative materials and the benefits and difficulties for students and teachers.

A next methodological topic which can be challenging in educational research is validity. In our studies, we tried to find balance between generaliza-

bility and ecological validity. We designed instructional units which were open to context modifications, and we entrusted the design of writing tasks to teachers themselves, accompanied by guiding sessions on design principles for effective writing-to-learn tasks. This procedure enhanced the feasibility of the design, and it provided a realistic view of how implementation would work in practice. We kept track of generalizability by consistently exploiting design principles. After implementation of a unit, we evaluated to what extent the design principles behind the design had been brought to bear. We considered the intervention to be a construct and defined and operationalized it as such (Rijlaarsdam et al., 2017). Teacher involvement was discussed in more depth as a third methodological issue. Although involving teachers in research creates engagement, it might be problematic from a researchers' perspective, since it also might impede a transparent research process.

Furthermore, we discussed three variables relevant in our studies, and how these variables were measured: content learning, epistemic experience, and writing proficiency. We decided to investigate students' content learning to explore whether we can promote writing development without hindering content knowledge gain in the history study. Although a recall test might not be the best type of measurement for content learning, it did illuminate us that knowledge accumulation was not limited by the strategy instruction. Next to that, students who received strategy instruction reported they experienced the epistemic function of writing, which might be seen as a confirmation of learning effects. Most importantly, students showed greater writing proficiency after strategy instruction.

3.3 Recommendations for Future Research and Educational Practice

The results of our research provide leads for future studies. We suggest a focus on theoretical understanding of disciplinary writing processes, writing strategy preferences, nuance differences between disciplines, and the relation between writing and learning across disciplines.

Additionally, this work resulted in several implications for practice, in three areas: writing task design, writing instruction and support, and teacher aids. Regarding writing task design, we recommend content teachers to include writing-to-learn tasks into their lessons; to have students write based on sources, which are deliberately selected; and to use short writing tasks, to be able to have students write within teachers' sight. Regarding writing instruction and support, we recommend teachers to provide and model strategies; to highlight disciplinary aspects in their process support; to support students'

writing during task performance; to use exemplars to start a discussion of text quality; and to connect product feedback to the student's process. Lastly, we recommend teacher educators to develop teachers' pedagogical content knowledge of writing; and we call on textbook publishers to develop teaching aids which include writing support.

4. CONCLUSION

We conclude that secondary school teachers who include disciplinary reading-writing tasks into their curriculum, and who provide additional support to students' writing process by means of strategy instruction, support the development of students' disciplinary literacy skills. Our studies also provide insights in teachers' interaction with design principles for writing task design and show what we can achieve if we involve teachers in designing and implementing interventions in general.

SAMENVATTING

Het doel van dit proefschrift was om bij te dragen aan de ontwikkeling van vakspecifieke taalvaardigheid van middelbare scholieren, met een focus op schrijfvaardigheid. Het wordt steeds urgenter om te werken aan de geletterdheid van leerlingen; volgens internationaal vergelijkingsonderzoek neemt de taalvaardigheid van leerlingen af, terwijl we leven in een steeds taliger maatschappij. Met dit promotieonderzoek onderzochten we hoe docenten van twee vakken, geschiedenis en filosofie, binnen hun domein zouden kunnen werken aan de disciplinaire geletterdheid van hun leerlingen, zonder daarbij de dagelijkse onderwijspraktijk al te veel te verstoren.

1. INLEIDING

In verschillende schoolvakken wordt schrijven over het algemeen gebruikt als middel om inhoud te verwerken of om de kennis en redeneringen van leerlingen te beoordelen. Er wordt weinig nadruk gelegd op leren schrijven. Dit ligt misschien voor de hand: leerlingen leren hoe ze een goede tekst moeten schrijven bij het vak Nederlands. Leren schrijven wordt over het algemeen gezien als de verantwoordelijkheid van de docent Nederlands.

Toch is het belangrijk om lees- en schrijfvaardigheidsactiviteiten in te bedenken in vaklessen, omdat er verschillen zijn tussen vakgebieden in hoe kennis wordt gecreëerd, welke strategieën worden gebruikt voor onderzoek, en welke genres, welke concepten en welk taalgebruik erbij betrokken zijn (Goldman, et al., 2016). Het is de verantwoordelijkheid van de vakdocent om deze vakspecifieke aspecten van taalvaardigheid te adresseren.

Uit eerder onderzoek is gebleken dat schrijfinstructie kan bijdragen aan het verbeteren van de schrijfkwaliteit en kennisverwerving in vaklessen (De La Paz, 2005; Foxworth & Mason, 2017). Vakdocenten beschikken echter niet altijd over de kennis om schrijfinstructie of -ondersteuning te bieden zoals docenten Nederlands die hebben (Gillis, 2014; Moje, 2008). Dat staat integratie van schrijfonderwijs in de vakles in de weg.

In dit proefschrift beoogden we dan ook om docenten in de bovenbouw van het voortgezet onderwijs te faciliteren in de ontwikkeling en ondersteuning van vakspecifiek schrijven. Dit deden we door nieuw lesmateriaal te ontwerpen, implementeren en testen. Omdat de schoolvakken geschiedenis en

filosofie beide hoge eisen stellen aan geletterdheid, besloten we ons op deze twee vakken te richten. Onze belangrijkste onderzoeksvraag was:

Welke aanpak is geschikt om de vakspecifieke taalvaardigheid van leerlingen effectief en efficiënt te ontwikkelen in de bovenbouw van het geschiedenis- en filosofieonderwijs?

We onderzochten deze onderzoeksvraag door middel van ontwerponderzoek: na een exploratie van de context en de literatuur, formuleerden we ontwerpprincipes voor schrijftaken en strategie-instructie. De taakprincipes hielpen docenten op weg bij het ontwikkelen van goede schrijf-leertaken die ze in hun lessen konden integreren. De strategie-instructieprincipes vormden de basis voor een lessenreeks, die steeds werd toegespitst op het betreffende vak. De lessenreeksen werden in verschillende studies getest in de lespraktijk.

2. OVERZICHT VAN DE ONDERZOEKEN

We begonnen met een voorstudie (beschreven in hoofdstuk 3). Het doel hiervan was om een duidelijk beeld te krijgen van de schrijfpraktijk van docenten geschiedenis en filosofie, en van hun behoeften aan verbetering van het vakspecifiek schrijven. We hielden interviews met docenten geschiedenis ($n = 10$) en filosofie ($n = 11$) van verschillende scholen. We vroegen hen welk(e) type(n) schrijftaken ze vaak gebruikten in hun vak in de bovenbouw. Docenten werd gevraagd een representatief voorbeeld van een schrijftaak voor hun vak mee te nemen, en we gebruikten deze voorbeeldtaak als "kapstok". We vroegen de docenten welke eisen ze stelden aan de uiteindelijke tekst, hoe ze het schrijf-proces van de leerlingen ondersteunden en hoe ze dachten dat het proces van de leerlingen eruit zou moeten zien.

Uit het vooronderzoek bleek dat docenten moeite hadden met het verwoorden van schrijfprocessen. Zowel filosofiedocenten als geschiedenisdocenten gaven aan dat ze het een uitdaging vonden om te beschrijven wat leerlingen eigenlijk zouden moeten doen om hun opdracht met succes af te ronden. De meerderheid van de geïnterviewde geschiedenisleraren noemde geen aspecten van het schrijfproces; ze richtten zich op het lezen en evalueren van bronnen, maar niet op bijvoorbeeld het structureren of reviseren van de geschreven tekst. Docenten filosofie waren zich meer bewust van de rol van schrijfprocessen in hun opdrachten; dit impliceert dat ze schrijven zagen als een middel om filosofie te leren.

2.1 Hoofdstuk 2

De volgende studie was ook een achtergrondstudie, maar dan gericht op leerlingen. We bestudeerden de processen van leerlingen uit 5 vwo, terwijl ze schrijftaken uitvoerden voor de twee betreffende vakken in een hardop-denkstudie. We selecteerden vijftien uitblinkers die beide vakken volgden, omdat we wilden onderzoeken welke cognitieve processen "beginnende ex-perts" zouden laten zien wanneer ze in de twee vakken bronteksten lazen en een tekst schreven. Voor elk vak selecteerden we een representatieve schrijftaak op basis van bronnen. Leerlingen voerden de twee taken in willekeurige volgorde uit terwijl ze hardop nadachten. We analyseerden hun denkpatronen om te zien of en in welke mate deze per vakgebied verschilden en of de cognitieve activiteiten de tekst- en denkkwaliteit konden verklaren.

Om de cognitieve processen van leerlingen te identificeren en patronen te ontdekken, gebruikten we een datagestuurde ("bottom-up") aanpak. Voor het segmenteren en coderen van de hardopdenkprotocollen van leerlingen maakten we ook gebruik van bestaande raamwerken voor lezen en schrijven in het algemeen (bijv. Hayes & Flower, 1981), en lezen en schrijven in de disciplines (bijv. Brante & Strømsø, 2018; Corcelles & Castelló, 2015; Hof et al., 2015). Drie hoofdcategorieën: lezen, schrijven en metacognitie, werden onderverdeeld in elf procesactiviteiten. Alle procesactiviteiten werden geanalyseerd op frequentie, en absolute en relatieve duur.

Vervolgens werden de protocollen van de leerlingen beoordeeld op twee maten van kwaliteit: (1) kwaliteit van de uiteindelijke tekst en (2) kwaliteit van het denkproces van de leerlingen. Daarna werden kwantitatieve analyses uitgevoerd: we onderzochten correlaties tussen de kwaliteit van de teksten van leerlingen, de kwaliteit van het denkproces van leerlingen en hun procesactiviteiten.

De resultaten toonden aan dat de relaties tussen kwaliteit en activiteiten verschilden per taak, of, zoals we suggereerden, per vakgebied. In de geschiedenisopdracht hing de tekstkwaliteit vooral af van de planningsvariabelen. Leerlingen die meer tijd besteedden aan planning en vaker planden, produceerden teksten van hogere kwaliteit. In de filosofieopdracht daarentegen waren de meeste variabelen die de kwaliteit beïnvloedden schrijfvariabelen. Leerlingen die teksten van relatief hoge kwaliteit schreven, vertoonden kenmerken van expert-schrijvers: ze produceerden langere teksten, besteedden meer tijd aan schrijven en reviseerden vaker.

Bovendien verschilde de relatie tussen tekstkwaliteit en denkproceskwaliteit per vak. De geschiedenisopdracht lokte vooral de uitkomst van het

redeneerproces van leerlingen uit, terwijl de filosofietaak leerlingen uitnodigde om ook hun hele redenering op te schrijven; de filosofietaak stimuleerde dat denken en schrijven samen opgingen.

De resultaten van deze hardopdenkstudie hadden implicaties voor de volgende stappen in ons onderzoek: we beschouwden het karakter van de filosofietaak, die het denkproces van leerlingen stimuleerde, als een geschikt taaktype voor gebruik in andere vaklessen. In tegenstelling tot de geschiedenisopdracht, die alleen een resultaat vereiste, kon de filosofietaak worden gekarakteriseerd als een schrijven-om-te-leren taak; zulke taken zijn zeer geschikt voor gebruik in inhoudslessen, omdat ze passen bij de focus op kennisontwikkeling.

2.2 Hoofdstuk 3

In een volgende studie werden ontwerpprincipes en een lessenreeks ontwikkeld om het historisch schrijven van leerlingen te bevorderen. De lessenreeks werd geëvalueerd op validiteit, bruikbaarheid en effectiviteit. Inzichten uit onze voorstudie, de hardopdenkstudie (hoofdstuk 2) en een literatuuronderzoek leidden tot de formulering van twee ontwerpprincipes, die de basis vormden voor de lessenreeks voor zowel geschiedenis als filosofie. De twee ontwerpprincipes waren:

Als we willen dat leerlingen een diepgaand begrip van de inhoud ontwikkelen door te schrijven, dan...

Ontwerpprincipe #1: dienen leerlingen korte evaluatieve teksten te schrijven, gebaseerd op meerdere primaire bronnen die meerdere perspectieven vertegenwoordigen.

Ontwerpprincipe #2: dienen leerlingen vakspecifieke, flexibele lees-schrijfstrategie-instructie te krijgen die gemakkelijk toepasbaar is voor docenten.

Strategie-instructie werd opgenomen in de lessenreeks, omdat kennisopbouw kan worden bespoedigd als het schrijfproces hanteerbaar gemaakt wordt. Flexibiliteit was van belang om tegemoet te komen aan verschillende schrijfprocesvoorkeuren. Een keuze aanbieden in de te nemen route kan voorkomen dat het schrijfproces een last wordt.

Als uitwerking van het tweede ontwerpprincipe, ontwierpen we een strategie op basis van de inzichten uit onze hardopdenkstudie en eerdere inter-

ventiestudies met schrijfstrategie-instructie in de vakles (bijv. De La Paz et al., 2017). Dit resulteerde in de uit zeven stappen bestaande Lees-Denk-Schrijfstrategie. Deze was toegespitst op historische evaluatieve vragen, in ons onderzoek "in hoeverre"-vragen genoemd. Met dit type vragen wilden we stimuleren dat schrijven het denken bevordert.

De strategie bood twee routes om tegemoet te komen aan verschillende voorkeuren voor het schrijfproces: "vrij schrijven", voor leerlingen die liever vanuit flow schrijven, om daarna flink te schaven aan de tekst, en "vooraf plannen", voor leerlingen die liever vanuit een tekstschema schrijven.

In een eerste testronde ontwierpen twee docenten geschiedenis drie schrijftaken en implementeerden zij de taken en de strategie-instructie in hun lessen in 5 vwo. We gebruikte een "switching replications design" met drie meetmomenten, om te kunnen onderzoeken of de lessenreeks effectief leek.

De teksten van de leerlingen werden beoordeeld door een jury van geschiedenisdocenten op drie criteria: holistische kwaliteit, inhoudelijke kwaliteit en kwaliteit van de structuur. De resultaten gaven blijk van effectiviteit op de inhoudelijke kwaliteit. De resultaten waren echter niet volledig overtuigend, mogelijk omdat het tweede principe niet voldoende uit de verf was gekomen in de lessen. De deelnemende docenten waardeerden de schrijftaken zoals beschreven in het eerste principe. Ze waren echter nog niet overtuigd van het nut van de strategie-instructie; de rationale achter deze schrijfinstructie was voor hen niet voldoende duidelijk. De belangrijkste aanbeveling uit dit onderzoek was daarom om de achterliggende gedachte achter het materiaal aan docenten uit te leggen, voordat zij het gaan gebruiken. In een herontwerp werd om deze reden een scholingssessie opgenomen.

2.3 Hoofdstuk 4

In een derde onderzoek hebben we de effectiviteit van de geoptimaliseerde lessenreeks getest in een quasi-experimenteel onderzoek met drie onderzoekscondities. We beoogden inzicht te krijgen in effecten van twee factoren: (1) implementatie van schrijftaken, en (2) implementatie van schrijfprocesinstructie.

In een eerste experimentele conditie (7 groepen, 119 leerlingen, 6 docenten) ontwierpen de deelnemende docenten schrijftaken die aansloten bij ontwerp-principe 1. Vervolgens implementeerden ze de geoptimaliseerde schrijfstrategie-instructie. In een tweede experimentele conditie (3 groepen, 63 leerlingen, 2 docenten) ontwierpen docenten ook schrijftaken, maar werd de strategie-

instructie weggelaten. Een derde conditie was een controleconditie waarin niet werd geschreven (4 groepen, 86 leerlingen, 3 docenten).

In een quasi-experimenteel design, met schrijftaken als voor- en nameting, wilden we effecten meten op de schrijfvaardigheid van leerlingen. Deze schrijftaken waren evaluatieve taken over een onderwerp dat niet besproken werd tijdens de geschiedenislessen. Na de strategie-instructie – die bestond uit directe instructie, modeling door middel van een video met modeling peers, en discussies over voorbeelden – oefenden de leerlingen het schrijven door het uitvoeren van twee schrijftaken, die waren ontworpen door hun docent en daarom volledig waren afgestemd op de inhoud van hun geschiedenislessen. Deze schrijf-leertaken verschilden dus van groep tot groep.

Een juryteam van geschiedenisdocenten beoordeelde de teksten van leerlingen op vijf aspecten: situationeel begrip, multiperspectiviteit, brongebruik, argumentatie en structuur. Bovendien beoordeelden ze de teksten holistisch, door middel van een tekstschaal. Deze scores waren indicatoren voor de schrijfvaardigheid van leerlingen, specifiek bij geschiedenis.

Verder onderzochten we de effecten van de schrijftaken en strategie-instructie op de "leerervaring" van leerlingen: in hoeverre hadden leerlingen de schrijftaken ervaren als leertaken? Tot slot onderzochten we of schrijftaken een goed alternatief waren voor andere, meer gebruikelijke leeractiviteiten voor het leren van geschiedenis. Daarom hebben we de vakinhoudelijke kennis van de leerlingen gemeten in een "recall test". Het onderwerp van deze test verschilde van klas tot klas, omdat elke docent zijn eigen lesprogramma volgde.

De interventie had effect op de schrijfvaardigheid van leerlingen: leerlingen die schrijfstrategie-instructie hadden gekregen, schreven betere teksten dan leerlingen die deze instructie niet hadden gehad. De mate van zelfvertrouwen bij schrijftaken en attitude ten opzichte van schrijven deden er hierbij niet toe. Significante effecten van de strategie-instructie waren zichtbaar in alle aspecten van tekstkwaliteit, holistisch en op vakspecifieke criteria. Bovendien gaven leerlingen die strategie-instructie kregen vaker dan leerlingen in de controleconditie aan dat ze de leerfunctie van schrijven hadden ervaren.

De resultaten van de recall test toonden aan dat leerlingen die strategie-instructie kregen even goed presteerden als leerlingen in de andere condities wat betreft de toename van vakinhoudelijke kennis. Deze resultaten geven aan dat schrijven een veelbelovende alternatieve activiteit lijkt voor het leren van geschiedenis. Dit leidt echter alleen tot betere teksten, als er ook aandacht wordt besteed aan het schrijfproces.

2.4 Hoofdstuk 5

Om verschillende disciplines met elkaar te kunnen vergelijken, onderzochten we in een vierde studie de effecten van een vergelijkbaar ontwerp voor filosofie. Dit onderzoek richtte zich op de interactie van drie filosofieleraars met het schrijftaakontwerp en de strategie-instructie die ze implementeerden in hun filosofielessen in 4 vwo. Het ontwerp voor geschiedenis werd aangepast aan filosofisch schrijven. Het ontwerp was niet dichtgetimmerd: er was ruimte voor de docent om het ontwerp aan te passen aan de eigen context. Om docenten aan te zetten tot reflectie over schrijfinstructie en tekstkwaliteit, voegden we in deze studie een begeleidend programma toe.

In deze ontwerpstudie kozen we voor een kwalitatieve insteek, om diepgaand te onderzoeken hoe filosofiedocenten omgingen met de ontworpen materialen en welke resultaten dit opleverde voor de vakspecifieke schrijfvaardigheid van leerlingen.

Drie filosofieleraars ontwierpen hun eigen schrijftaken en implementeerden de ontworpen strategie-instructie. We volgden en evalueerden de implementatie om te onderzoeken hoe veerkrachtig het ontwerp zou zijn in open, ecologisch valide contexten en we hielden evaluatieve interviews met leerlingen en docenten om hun aanpassingen aan het ontwerp te onderzoeken.

Bovendien onderzochten we de vooruitgang van leerlingen. De schrijfvaardigheid van de leerlingen werd gemeten met drie schrijftaken, waarvan er twee werden ontworpen door de docenten van de klas. Deze metingen verschilden dus van klas tot klas.

Docenten selecteerden vervolgens een benchmark voor elke taak: de "gemiddelde" tekst voor die taak. Deze activiteit zette aan tot nadenken over kwaliteitscriteria in filosofisch schrijven en over vooruitgang van leerlingen.

De schrijftaken en de teksten van de leerlingen werden beoordeeld door onafhankelijke juryteams bestaande uit docenten filosofie. De jury beoordeelde de teksten holistisch, met steeds de door de docent gekozen benchmark als referentiepunt. In reflectieve interviews presenteerden we de scores aan de vakdocenten, en vroegen hen om hierop te reflecteren.

De resultaten lieten zien dat filosofiedocenten goed uit de voeten konden met de theoretisch onderbouwde ontwerpprincipes voor schrijftaken. Wel hadden ze moeite met de lengte van de taken; typische filosofietaken zijn doorgaans langer en er wordt meer tijd gegeven om te schrijven. De strategie-instructie bleek haalbaar en leerzaam voor leerlingen.

Met betrekking tot de vooruitgang van leerlingen concludeerden we voorzichtig dat de lessenreeks bijdroeg aan de filosofische schrijfvaardigheid van

leerlingen. Leerlingen lieten na de interventie meer zelfstandig filosofisch denken zien in hun tekst, terwijl de taken complexer werden.

Het begeleidingsprogramma van de docenten ontlokte reflectie op de waarde van schrijftaken en -instructie voor de vakdoelen, en op de perceptie van de vooruitgang van de leerlingen.

3. DISCUSSIE

3.1 Conceptuele kwesties

In de vier studies van dit proefschrift zijn drie ontwikkelingslijnen naar voren gekomen, namelijk met betrekking tot (1) vakspecifiek schrijven; (2) schrijfinstructie; en (3) verbetering van de lespraktijk.

In dit proefschrift pleiten we voor de integratie van schrijftaken die het leren kunnen ondersteunen, om hiermee te werken aan de vakspecifieke taalvaardigheid van leerlingen. Leren wordt over het algemeen bevorderd wanneer leerlingen worden aangemoedigd om verschillende perspectieven op een kwestie te verkennen, door meerdere bronnen te analyseren en interpreteren, en door die te verwerken in een evaluatieve tekst.

De resultaten van onze studies toonden aan dat docenten de evaluatieve schrijftaken waardeerden als leeractiviteiten, docenten geschiedenis in het bijzonder. Docenten filosofie hadden moeite met de tijdsduur die leerlingen hadden om een schrijftaak te maken: die was beperkt tot één lesuur. Hoewel deze keuze indruiste tegen hun opvatting dat filosofisch schrijven eigenlijk meer tijd vergt, erkenden de filosofiedocenten ook dat korte taken een middel kunnen zijn om de procedure van filosofisch schrijven onder de knie te krijgen.

3.2 Methodologische kwesties

We bespraken een aantal methodologische kwesties. Zo concludeerden we dat de combinatie van verschillende onderzoeksmethoden een kracht was; het heeft verschillende perspectieven belicht. Kwantitatieve metingen lieten zien dat schrijfonderwijs effectief is voor de ontwikkeling van de schrijfvaardigheid en inhoudelijke kennis van leerlingen. Kwalitatieve methoden gaven ons inzicht in de ervaringen van docenten bij het implementeren van de materialen, en de voordelen en moeilijkheden voor leerlingen en docenten.

Een methodologische kwestie die een uitdaging kan vormen in onderwijsonderzoek is validiteit. In onze studies probeerden we een balans te vinden tussen generaliseerbaarheid en ecologische validiteit. We ontwierpen instruc-

tie die open stond voor contextaanpassingen en we vertrouwden het ontwerp van schrijftaken toe aan docenten zelf. Deze procedure verbeterde de haalbaarheid van het ontwerp en gaf een realistisch beeld van hoe de implementatie in de praktijk zou werken. We hielden de generaliseerbaarheid in de gaten door consequent gebruik te maken van ontwerpprincipes. Na implementatie van een eenheid evalueerden we in hoeverre de ontwerpprincipes achter het ontwerp naar voren waren gekomen. We beschouwden de interventie als een construct en definieerden en operationaliseerden het ook als zodanig, het advies van Rijlaarsdam et al. (2017) volgend.

Een derde methodologische kwestie is de rol van docenten bij het doen van onderzoek. Voor het uiteindelijke praktijkdoel is het betrekken van docenten in het onderzoek gunstig, maar vanuit het perspectief van onderzoekers kan die betrokkenheid ook problematisch zijn, omdat het een transparant onderzoeksproces in de weg kan staan. Waar docenten hun eigen weg gaan, ontstaat immers enige ruis.

In onze studies zijn drie variabelen steeds teruggekomen: inhoudelijk leren, de mate van leerervaring, en schrijfvaardigheid van leerlingen. De belangrijkste bevinding is dat leerlingen betere vakteksten schreven na strategie-instructie; dat was ons voornaamste doel. We besloten daarnaast het inhoudelijk leren van leerlingen te onderzoeken om na te gaan of we de schrijfvaardigheid van leerlingen konden bevorderen zonder de toename van inhoudelijke kennis te belemmeren. Hoewel een recall test misschien niet het beste type meting is voor inhoudelijk leren, liet het ons zien dat de kennistoename niet beperkt werd door de strategie-instructie. Als het gaat om de leerfunctie van schrijven, rapporteerden leerlingen die strategie-instructie kregen dat ze de leerfunctie van schrijven hadden ervaren. Dat kan gezien worden als een bevestiging van leereffecten.

3.3 Aanbevelingen voor toekomstig onderzoek en onderwijspraktijk

De resultaten van ons onderzoek bieden aanknopingspunten voor toekomstig onderzoek. Te denken valt aan studies gericht op theoretisch begrip van disciplinaire schrijfprocessen, voorkeuren van leerlingen voor de ene of de schrijfstrategie, nuanceverschillen tussen vakgebieden en de relatie tussen schrijven en leren in verschillende vakgebieden.

Daarnaast resulteerde dit werk in verschillende implicaties voor de praktijk, op drie gebieden: het ontwerp van schrijfofdrachten, schrijfinstructie en -ondersteuning, en hulpmiddelen voor docenten. Wat betreft het ontwerp van schrijfofdrachten raden we vakdocenten aan om schrijf-leeropdrachten in

hun lessen op te nemen, die aansluiten bij vakinhoudelijke doelen; om leerlingen te laten schrijven op basis van goed geselecteerde bronnen; en om korte schrijftaken te gebruiken, zodat leerlingen binnen het zicht van de docent kunnen schrijven.

Wat schrijfinstructie en -ondersteuning betreft, raden we docenten aan om strategieën aan te reiken en die te modelleren; om vakspecifieke aspecten te benadrukken in hun procesondersteuning; om het schrijfproces van leerlingen te ondersteunen tijdens het uitvoeren van de taak; om voorbeeldteksten te gebruiken om een discussie over tekstkwaliteit op gang te brengen; en om productfeedback te koppelen aan het proces van de leerling.

Tot slot raden we lerarenopleiders aan om de pedagogisch-didactische kennis van schrijven bij docenten te ontwikkelen en roepen we uitgevers van studieboeken op om hulpmiddelen te ontwikkelen die docenten kunnen inzetten bij het begeleiden van schrijftaken in de les.

4. CONCLUSIE

We concluderen dat docenten in het voortgezet onderwijs die disciplinaire lees-schrijftaken opnemen in hun curriculum en die het schrijfproces van leerlingen extra ondersteunen door middel van strategie-instructie, de ontwikkeling van disciplinaire geletterdheid van leerlingen bevorderen. Onze studies geven ook inzicht in de interactie van docenten met ontwerpprincipes voor het ontwerpen van schrijftaken en laten zien wat we kunnen bereiken als we docenten betrekken bij het ontwerpen en implementeren van interventies in het algemeen.

CONTRIBUTIONS OF AUTHORS

Chapter 2

Lieke Holdinga: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Writing: original draft, Writing: review and editing.

Tanja Janssen: Conceptualization, Methodology, Supervision, Writing: review and editing.

Gert Rijlaarsdam: Conceptualization, Methodology, Supervision, Writing: review and editing.

Chapter 3

Lieke Holdinga: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Writing: original draft, Writing: review and editing.

Jannet van Drie: Methodology, Supervision, Writing: review and editing.

Tanja Janssen: Conceptualization, Methodology, Supervision

Gert Rijlaarsdam: Conceptualization, Methodology, Supervision, Writing: review and editing.

Chapter 4

Lieke Holdinga: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Writing: original draft, Writing: review and editing.

Jannet van Drie: Conceptualization, Methodology, Supervision, Writing: review and editing.

Gert Rijlaarsdam: Conceptualization, Formal analysis, Methodology, Supervision, Writing: review and editing.

Chapter 5

Lieke Holdinga: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Writing: original draft, Writing: review and editing.

Jannet van Drie: Conceptualization, Methodology, Supervision, Writing: review and editing.

Gert Rijlaarsdam: Conceptualization, Methodology, Supervision, Writing: review and editing.

PUBLICATIONS AND PRESENTATIONS

Scientific Publications

- Holdinga, C. C., Van Drie, J. P., & Rijlaarsdam, G. C. W. (nd). Fostering philosophy teachers' disciplinary writing practice; A multiple-case design study. Under review.
- Holdinga, C. C., Van Drie, J. P., & Rijlaarsdam, G. C. W. (nd). Underrated yet effective: writing process instruction in the history classroom. Under review.
- Holdinga, C. C., Van Drie, J. P., Janssen, T. M., Rijlaarsdam, G. C. W. (2023). Writing to learn history: An instructional design study. *L1-Educational Studies in Language and Literature*, 23(1), 1-44.
<https://doi.org/10.21248/l1esll.2023.23.1.526>
- Flores-Ferrés, M., van Weijen, D., van Ockenburg, L., ten Peze, A., Alkema, E., Holdinga, C., & Rijlaarsdam, G. (2023). Understanding and being understood: Three basic design principles for effective writing education. In: A. Galvão Spinillo, & C. Sotomayor (Vol. Eds.), *Development of writing skills in children in diverse cultural contexts*. Springer.
- Holdinga, C. C., Janssen, T. M., & Rijlaarsdam, G. C. W. (2021). The relationship between students' writing process, text quality, and thought process quality in 11th-grade history and philosophy assignments. *Written Communication*, 38(4), 544-586. <https://doi.org/10.1177/07410883211028853>

Professional Publications

- Holdinga, C. C. (2022). Wat is het belang van vakspecifieke taalvaardigheid? In WODN Werkgroep Onderzoek Didactiek Nederlands, Levende Talen (Ed.), *Handboek Didactiek Nederlands*. <https://didactiek-nederlands.nl/handboek/2022/04/wat-is-het-belang-van-vakspecifieke-taalvaardigheid/>
- Holdinga, C. C. (2022). Filosoferen bij Nederlands. In: G. Rijlaarsdam, M. Braaksmā, M. van de Guchte, A. ten Peze, M. Schrijvers, & D. van Weijen (Eds.), *Literatuur leren lezen in dialoog. Lessen geïnspireerd op het werk van Tanja Janssen*. Amsterdam: Amsterdam University Press.
- Holdinga, C. C., Janssen, T. M., Van Drie, J. P., & Rijlaarsdam, G. C. W. (2022). Het nut van opschrijven: de functies van schrijftaken bij geschiedenis en filosofie nader beschouwd. *Dimensies*, 4, 97-115.

Contributions to International Conferences

- Holdinga, C. C. (2022). *Goed redeneren, matig formuleren. Hoe leerlingen lees-schrijftaken maken bij geschiedenis en filosofie* [Paper presentation]. Onderwijs Research Dagen, Hasselt, Belgium.
- Luna, M., Holdinga, C., Rijlaarsdam, G., Van Drie, J., Villalon, R., Martín, E., Núñez, J. A., Martínez-Álvarez, I., & Cuevas, I. (2022, June 20-22). *Supporting teachers to develop interventions on argumentative writing strategies* [Round table session]. SIG 12 Writing Biennial Conference, Umea, Sweden.
- Holdinga, C. C., Janssen, T. M., Van Drie, J. P., & Rijlaarsdam, G. C. W. (2022, June 20-22). *Learning history through evaluative writing tasks: an instructional design study* [Paper presentation]. SIG 12 Writing Biennial Conference, Umea, Sweden.
- Holdinga, C. C., Janssen, T. M., & Rijlaarsdam, G. C. W. (2021, August 23-27). *Secondary school students' cognitive processes while writing-from-sources in history and philosophy* [Paper presentation]. EARLI 2021, Gothenburg, Sweden.
- Holdinga, C. C., Janssen, T. M., & Rijlaarsdam, G. C. W. (2019, June 26-28). *Writing from sources in history and philosophy courses: Teachers' practices and beliefs*. [Poster presentation]. ARLE 2019, Lisbon, Portugal.

Contributions to Dutch Conferences

- Holdinga, C. C. (2022, December 2). *Bundel de krachten in de school: een 'good practice' voor schrijfonderwijs* [Workshop]. Levende Talen Congres Online.
- Holdinga, C. C. (2022, November 10). *Beter schrijven bij geschiedenis en filosofie* [Workshop]. Werkconferentie SPRONG & Platform Taalgericht Vakonderwijs, Tilburg.
- Holdinga, C. C. (2022, March 28). *Schrijven bij filosofie* [Workshop]. Landelijke Lio-dag Filosofie, Utrecht.
- Holdinga, C., & Alkema, E. (2021, October 5). *Schrijven bij andere vakken* [Workshop]. VONK Inspiratiemiddag 2021.

ABOUT THE AUTHOR

Lieke Holdinga (De Bilt, 1985) obtained her master's degree in Language Development cum laude from Utrecht University in 2007. She then enrolled the teacher training program at IVLOS (Utrecht University) and obtained her teaching degree in 2008. Since 2008 she has been teaching Dutch language and literature at Vituscollege in Bussum.

In 2012 Lieke started the two-year post-master trajectory Academic Mastery. This research master's program prepares teachers to do research in education. As a thesis for this program, Lieke studied writing skills in other subjects. In 2014, she obtained the master's degree in educational science.

The aforementioned thesis resulted in a research proposal for a doctoral dissertation entitled "Learning-to-write and writing-to-learn in history and philosophy", for which Lieke was assigned a Dudoc-Alfa grant in 2018. The Dudoc-Alfa organization offers teachers the opportunity to conduct four years of research on teaching practice. This resulted in the current dissertation. In the meantime, she presented her research in international (ARLE, 2019; SIG Writing, 2022), and national settings (Levende Talen, 2022).

Together with fellow teacher-researchers, she started VONK, an online conference for teachers of Dutch language and literature. The goal of this initiative was to stimulate dissemination of valuable educational research insights into practice.

Currently Lieke works as a teacher of Dutch language and literature at Vituscollege Bussum. Furthermore, she is an active member of the Werkgroep Mondelinge Taalvaardigheid (WMT) of Levende Talen Nederlands. The WMT-group contains experts on speaking and conversation skills, who aim to compile and share knowledge, skills and attitudes related to speaking, conversation and listening with education and related organizations. Recently, Lieke became a member of the editorial team for the digital handbook of Dutch Language Teaching, which was initiated by the Werkgroep Onderzoek en Didactiek Nederlands (WODN) (<https://didactieknederlands.nl/handboek/>).

DANKWOORD

Een promotieonderzoek doe je niet alleen. In de afgelopen jaren hebben veel mensen mij geholpen, die ik hier graag wil bedanken.

Als eerste natuurlijk mijn promotor, Gert Rijlaarsdam. Je betrokkenheid is groots. Wat een geluk dat ik onder jouw supervisie kon starten met het onderzoek. Ik heb heel veel waardering voor hoe je steevast klaar hebt gestaan om me van advies te voorzien als ik dat nodig had. Je scherpe blik en het tempo waarin je werkt zijn absoluut bewonderenswaardig.

De tweede persoon aan wie ik dank wil uitspreken, is Tanja Janssen. Je enthousiasmeerde me om een voorstel te schrijven voor een onderzoek en bracht me in contact met Gert. Je bent ontzettend scherpzinnig, maar er bleek ook altijd ruimte voor geruststellende woorden.

Later in het traject versterkte Jannet van Drie het team. Het is niet makkelijk om in een team dat zijn weg heeft gevonden, weer een nieuwe modus te vinden. Maar de nieuwe dynamiek heeft goed uitgediend en je bent met je perspectief vanuit de geschiedenisdidactiek een heel welkome aanvulling geweest op het juiste moment.

Mijn grote dank gaat ook uit naar de stuurgroep van Dudoc-Alfa. De commissie gunde mij een beurs en ik ben ontzettend dankbaar voor het vertrouwen dat ik kreeg om deze beurs goed te besteden.

Dank ben ik ook verschuldigd ook aan de schoolleiding van het Vituscollege, voor de ruimte die mij is gegeven om te kunnen doen wat ik leuk vind. Daarnaast was er altijd een mogelijkheid om de kennis die ik opdeed te delen met collega's, in het bijzonder de collega's van mijn eigen sectie Nederlands. Speciale dank spreek ik uit aan René van Bommel, die mij in 2008 de vrijheid gaf om te beginnen met het onderzoek, omdat hij zag dat ik dit gewoonweg heel graag wilde doen. Je oog voor het welzijn van mensen is een grote kracht.

Omdat ik graag verbetering wilde bewerkstelligen in het onderwijs, en dat uiteraard ook graag wilde op mijn eigen school, had ik hulp nodig van collega's. Gelukkig waren er veel collega's die geïnteresseerd waren om mee te denken en te doen. Heel veel dank aan mijn altijd nieuwsgierige collega Inno voor het mee-ontwikkelen en testen van het lesmateriaal voor geschiedenis.

Een heel aantal docenten van andere scholen dan mijn eigen meldde zich vrijwillig aan voor deelname. Het was heel inspirerend om met deze enthousiaste groep mensen te werken. Wat fijn dat er zo veel mensen betrokken zijn

bij het onderwijs, die echt hun best doen om goed les te geven en het onderwijs voor leerlingen blijven verbeteren. Natuurlijk dank ik ook de leerlingen die hebben willen meedoen met de lessen die ik ontwikkelde, en de leerlingen die meewerkten aan de video's die we hebben ingezet als lesmateriaal.

Mijn dank gaat daarnaast uit naar alle beoordelaars op wie ik een beroep deed voor de data-analyse. Stapels teksten om te beoordelen. Dan weer met een rubric, dan weer met Comproved, dan weer met een schaal. Speciale dank aan Arie en Michael, die nooit "nee" zeiden op mijn vraag.

Naast het begeleidingsteam en de hulptroepen bij het verzamelen van data, waren er collega's die op een andere manier bijdroegen. Graag bedank ik de Lunchlab-groep en de kamergenoten van kamer D8.12a: Anouk, Liselore, Edith, Magdalena, Thao, Chelsea, Marrit, Chris, Genja, Lucija, Stefan, Sevinç, Marloes, Monique en Floor. Er was altijd ruimte om te sparren tijdens een koffiepauze.

Speciale dank aan Daphne. Je zorgde voor een warm welkom in de Lunchlab-groep en in het creëren van de vertrouwde sfeer tijdens de lunchbijeenkomsten speelde jij zeker een sleutelrol. Dank ook voor je tip van een noisecancelling headphone, en dank aan BOSE voor het produceren van deze geweldige lifesaver.

Veel aanmoediging heb ik ook beleefd in het contact met andere promovendi van Dudoc-Alfa en NWO. In het bijzonder veel dank aan Gijs, Maartje, Renate, Johan, en de andere neerlandici van VONK.

Aan de zijlijn stonden daar altijd vrienden en familie. Dank aan ieder van jullie voor de interesse en warme betrokkenheid. Dank, Mechteld, voor het lezen en corrigeren van de teksten die ik schreef en natuurlijk de support. Speciale dank ook aan Patrick, voor het strakke design van het lesmateriaal en de cover van dit boek.

Mijn grote dank gaat ook naar mijn lieve ouders, die het pad naar de universiteit voor mij hebben vrijgemaakt, en daarmee dit promotietraject hebben mogelijk gemaakt. Jullie waren er altijd voor mij, en nog steeds.

Tot slot bedank ik mijn gezin voor alle liefde en flexibiliteit. Dank lieve Luc en Art, dat jullie ervoor zorgden ervoor dat ik de nodige afstand tot het werk bewaarde, omdat aandacht voor elkaar het allerbelangrijkste is als we bijeen zijn. Tot slot: Martin, mijn geweldige echtgenoot, dankjewel voor je eeuwige support. Wat de toekomst ook mag brengen: als het maar samen is.

DISSERTATION SERIES

- | | | |
|-------------------------------------|------|---|
| Jannemieke van de Gein (NL) | 1991 | The sense of sentence. Effects of grammar skills on aspects of writing proficiency |
| Rob Schoonen (NL) | 1991 | Constructie en validering van gestructureerde stelopdrachten voor zesdeklassers [Construction and validation of structured writing tasks for eight graders] |
| Amos van Gelderen (NL) | 1992 | Fonetisch en linguïstische validatie van oordelen over spreekprestaties [Phonetic and linguistic validation of rating of oral performances] |
| Michel Couzijn (NL) | 1995 | Observation of writing and reading activities: Effects on learning and transfer |
| Marianne Overmaat (NL) | 1996 | Empirische studies naar effectieve schrijfdidactiek [Empirical studies in written composition methodology] |
| Tanja Janssen (NL) | 1998 | Literatuuronderwijs bij benadering [Approaches in literature education] |
| Martine A.H. Braaksma (NL) | 2002 | Observational learning in argumentative writing |
| Hein Broekkamp (NL) | 2003 | Task demands and test expectations. Theory and empirical research on students' preparation for a teacher-made test |
| Isabelle de Ridder (BE) | 2003 | Hyperlinks & vocabulary learning from texts in L2 |
| Rita Rymenans & Sven de Maeyer (BE) | 2004 | Onderzoek naar kenmerken van effectieve scholen. Kritische factoren in een onderzoek naar schooleffectiviteit in het technisch en beroepssecundair onderwijs in Vlaanderen [Investigating characteristics of effective schools. Critical factors in a study of school effectiveness in technical and vocational secondary education in Flanders] |
| Lap Quoc Trinh (Vietnam) | 2005 | Stimulating learner autonomy in English language education. A curriculum innovation study in a Vietnamese context |
| Marleen Kieft (NL) | 2006 | The effects of adapting writing instruction to students' writing strategies. |
| Rita Rigo-Toth (Australia) | 2007 | A study of the influence of home and preschool factors on young children's literacy achievement |
| Ton Koet (NL) | 2007 | Polder English in Dutch ears. Empirical studies on the evaluation of the pronunciation of English as a foreign language |
| Theo Witte (NL) | 2008 | Het oog van de meester |
| Mariet Raedts (BE) | 2008 | De invloed van zelfeffectiviteitsverwachtingen, taakkennis en observerend leren bij een nieuwe en complexe schrijftaak [The effect of self-efficacy expectations, task knowledge and observational learning in a new and complex writing task] |
| Elke Van Steendam (BE) | 2008 | Effective instructional strategies in collaborative revision in ESL. The effects of two empirical studies |
| Daphne van Weijen (NL) | 2009 | Writing processes, text quality, and tasks effects. Empirical studies in first and second language writing |

Lieve Verheyden (BE)	2010	Achter de lijn. Vier empirische studies over ontluikende stelsvaardigheid [The story behind the line. Four empirical studies on writing by third and fourth graders of primary school]
Tatiana Berden-Antonenko (NL)	2010	Stimulating intercultural intellectual capabilities in intercultural communication. Testing an innovative course design
Talita Groenendijk (NL)	2012	Observe and explore. Empirical studies on learning in creative writing and visual arts.
Marion Tillema (NL)	2012	Writing in first and second language. Empirical studies on text quality and writing processes
Phuong Nam Thi Nguyen (Vietnam)	2012	Second language writing and literary reading in university: Three empirical studies
Dick van Dijk (NL)	2013	De taal van oplossingen. Een empirisch begrippenkader voor oplossingsgerichte interactie [The Language of Solutions. An empirical conceptual framework for solution-focused interaction]
Milan Kríz (Tsjechië)	2014	Teaching Dutch as a foreign language in Eastern Europe
Marco Kragten (NL)	2015	Comprehending process diagrams in biology education
Marrit Hoeks-van de Guchte (NL)	2015	Focus on form in task-based language teaching
Marie-Thérèse van de Kamp (NL)	2017	Re-imagine, re-design and transform. Enhancing generation and exploration in creative problem finding processes in visual arts education
Marloes Schrijvers (NL)	2019	The story, the self, the other. Developing insight into human nature in the literature classroom
Klaske Elving (NL)	2019	Effectieve leeractiviteiten voor het schrijfonderwijs in havo 4 [Effective learning activities for the teaching of writing in general secondary education, grade 10]
Suzanne Luger (NL)	2020	Lost in Latin translation. Teaching students to produce coherent target texts
Nina Vandermeulen (BE)	2020	Synthesis writing in upper-secondary education
Magdalena Flores-Ferrés (Chile)	2021	Teaching writing in Chile. An evidence base for policies on writing instruction for grades 7-12 of Chilean public schools
Martijn Koek (NL)	2022	Think twice: Literature lessons that matter
Liselore van Ockenburg (NL)	2022	Synthesis writing. Teaching high school students how to read, plan, draft, and revise
Edith Alkema (NL)	2022	Schrijven met kennis van zaken. Een didactiek voor de zaakvakken [Knowledgeable writing. Writing instruction for social sciences]
Chelsea O'Brien (NL)	2023	Dealing with setbacks. Exploring academic buoyancy in Latin students who are cognitively gifted