



## UvA-DARE (Digital Academic Repository)

### Talking about values : a dialogic approach to citizenship education as an integral part of history classes

Schuitema, J.A.

**Publication date**  
2008

[Link to publication](#)

#### **Citation for published version (APA):**

Schuitema, J. A. (2008). *Talking about values : a dialogic approach to citizenship education as an integral part of history classes*. [Thesis, fully internal, Universiteit van Amsterdam]. UvA/ILO.

#### **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.

## Chapter 4

### **TWO INSTRUCTIONAL DESIGNS FOR DIALOGIC CITIZENSHIP EDUCATION: AN EFFECT STUDY<sup>1</sup>**

The study presented in this chapter investigated the effectiveness of dialogic citizenship education in history classes. Two curriculum units were developed and implemented in the 8<sup>th</sup> grade of pre-university education. Both curriculum units aim to stimulate students' ability to take moral values and multiple perspectives into account when justifying their viewpoints. The two curriculum units differed in the balance between group work and whole-classroom teaching. The effects of both units were assessed. The ability to take values and multiple perspectives into account was assessed by means of short essays that students wrote about a moral issue. The results show a positive effect of group work. Students who worked relatively more often in small groups referred more to values and different perspectives in comparison to students who participated more often in whole-classroom teaching.

#### 1. INTRODUCTION

A renewed interest in citizenship education has emerged in the last two decades. The moral dimensions of civic participation and schools' role in preparing students for participation in society have in particular increasingly been emphasized. In contemporary discourses on citizenship education, citizenship implies more than an activity in the political sphere. It refers to a combination of knowledge, skills and attitudes for participation in the practices of society and engagement in public efforts to promote the common good (Althof & Berkowitz, 2006). The moral and social development of students is therefore considered essential to citizenship education (Althof & Berkowitz, 2006; Carr, 2006; Haste, 2004; Veugelers, 2007).

Despite this renewed interest in citizenship education and the moral and social development of students, relatively little is known about the effectiveness of the various teaching methods aimed at the social and moral development of students (Schuitema, Ten Dam & Veugelers, in press; Solomon, Watson & Battistich, 2001).

---

<sup>1</sup> Schuitema, J. A., Veugelers, W., Rijlaarsdam, G., & ten Dam, G. (submitted). Two instructional designs for dialogic citizenship education: An effect study.

A possible reason for this is the distinction that is often made between care for the moral and social development of students on the one hand, and teaching domain-specific knowledge and skills on the other. Most attention to the normative aspect of teaching currently concerns the daily interactions and rules and regulations at school or special activities in school. Citizenship education is usually realized in civic-education classes or in extracurricular projects such as community service (Althof & Berkowitz, 2006; Schuitema et al., in press). Our position, however, is that acquiring domain-specific knowledge and skills cannot be separated from moral development. The knowledge and skills acquired should enable students to judge and act independently with regard to moral issues, so that they can play an active role in society. This implies that students should learn to reflect on values represented in the subject matter, on their personal relationship to these values, and on the social relevance of the subjects, which are by definition value-related (see also Yore, Bisanz & Hand, 2003; Sadler & Zeidler, 2005). This makes citizenship education an inextricable part of the various school subjects.

History is one of the disciplines where a lively discussion is currently in progress on the relation between history education and citizenship education (Wrenn, 1999; Phillips, 2002; Barton & Levstik, 2004). Every society feels the need to pronounce moral judgements on what took place in the past (Barton & Levstik, 2004). This applies to students as well; they usually have opinions on historical topics too. A discipline such as history could utilize these moral judgements by making students aware of their own values, and teaching them to substantiate their judgements. History education, therefore, provides good opportunities for citizenship education. However, research mainly carried out in the 1980s shows that there is generally little room in history classes for input from students and for learning activities that stimulate problem solving and critical thinking (Wilson, 2001).

The study presented in this chapter was aimed at acquiring insight into which instructional design is the most effective for citizenship education in history lessons. Before outlining the methodology of the study, we will first discuss what the aim of citizenship education should be and what the literature tells us about effective teaching methods.

### *1.1 Learning objectives of citizenship education*

The aim of citizenship education is to prepare students for participation in society. Constituent elements of present-day western society are democracy and pluriformity (Banks, 2004; Parker, 1997b). According to Gutmann and Thompson (2004) a democratic society is characterized by continuous debate about moral values such as justice and the public interest. Participation in society means that citizens may contribute to this debate; it means thinking and talking about issues concerning justice and the public interest (Althof & Berkowitz, 2006; Haste, 2004; Parker, 1997b). Citizenship education cannot, therefore, be limited to teaching knowledge about how democratic society functions, nor should it be limited to transferring certain standards and values. Citizenship education must be directed at the attitudes and skills that students need to reflect on moral considerations and be able to justify them

(Veugelers & Vedder, 2003). An important aim of citizenship education, therefore, is the enhancement of students' competences to develop personal viewpoints on value-related matters and to justify their opinions to others. The fact that democracy is about plurality and difference also makes demands on citizenship education. It is essential that students understand that there are multiple perspectives on moral and social issues and that their own view is only one of many possible perspectives (Banks, 2004).

What does this mean for educational practice? How can history lessons contribute to citizenship education? First of all, students need to be able to reflect on the moral values that are at stake and take them into account when justifying their viewpoints to others (Veugelers, 2000). Secondly, students need to take into account and reflect on multiple perspectives while developing their own point of view (Oser, 1986). Obviously, students should also be able to apply these skills outside the context in which they were acquired; they should be able to transfer their skills to new issues.

### *1.2 Teaching methods aimed at the moral and social development of students*

Solomon et al. (2001) make a distinction between *direct* and *indirect* approaches to moral education. Direct approaches aim to transfer values. The teacher defines the values, explicitly propagates them, and encourages their application through a system of reward and punishment. The teacher may function as an example and role model (Doyle, 1997; Elkind & Sweet, 1997). Indirect approaches involve the implementation of dialogues (Grant, 1996; Noddings, 1995; Parker, 1997a; Tappan, 1998). Dialogues supposedly help students to develop the skills and attitudes citizens need to participate in society, such as critical-thinking skills for solving moral dilemmas (Blatt & Kohlberg, 1975; Frijters, Ten Dam & Rijlaarsdam, in press; Ten Dam & Volman, 2004), communicative skills (Parker, 1997a), and attitudes like tolerance and respect for the opinions of other people (Grant, 1996; Saye, 1998). Often a problem-based instructional design is recommended in which students work independently and control their own learning processes to a certain extent (Beane, 2002; Schultz, Barr & Selman, 2001), and in which collaborative efforts are needed (Covell & Howe, 2001; Murray, 1999; Saye, 1998). Proponents of indirect approaches to moral education usually emphasize the importance of classroom climate and school culture. Students need to have sufficient confidence to express their opinions and feel free to do so, even if their opinions differ from those of other students and teachers. It is, therefore, essential that teachers and students treat each other with respect (Power, Higgins & Kohlberg, 1989), and that students are encouraged to give their opinion in the classroom (Covell & Howe, 2001).

After an extensive analysis of research on moral and social education, Solomon et al. (2001) concluded that there is more evidence for the effectiveness of the social and moral development of students through the indirect teaching methods than through the direct methods. More specifically, there are indications that indirect approaches, where students work in small groups and engage in dialogue with each other, have a positive effect on attitudes regarding the universal rights of man

(Covell & Howe, 2001), result in less racist attitudes (Schultz et al., 2001), and make students think beyond their own personal interest (McQuaide, Leinhardt & Stainton, 1999).

Research on the effectiveness of teaching methods for the moral and social development of students is scarce. Moreover, its quality leaves much to be desired (Schuitema et al., in press; Solomon et al., 2001). Many of the proposed instructional designs have not been sufficiently elaborated. Hardly any attention at all has been paid to the attitudes and skills that are necessary for learning through dialogue (see e.g. Van der Linden & Renshaw, 2004). Moreover, the quality of the research designs is often weak: pre-tests or a control group are often absent. Solomon et al. (2001) further conclude that studies, which do include a control group, often compare a complete programme with the absence of such a programme. This makes it difficult to determine the effectiveness of the separate elements of the programme in question and the most effective combination of elements.

### *1.3 Instructional design for dialogic citizenship education*

In line with the indirect approach to moral education, we expected that stimulating dialogue in the classroom would be an effective teaching method to enhance students' ability to take moral values and multiple perspectives into account. A dialogue engages students in the learning process, stimulates participation, makes them aware of their values, and confronts them with other people's views.

To realize the desired results, a dialogue between students must satisfy a number of requirements. A dialogue that facilitates learning should be aimed at reaching agreement or understanding the other person (see, e.g., Grant, 1996; Parker, 1997a). Participants should be open to each other's views, develop their own views further, and cope with the different perspectives in a constructive way. Research shows that a dialogue aimed at generating ideas communally has a positive effect on students' reasoning capacities and their ability to reflect on values (Berkowitz & Gibbs, 1983; Frijters et al., in press; Rojas-Drummond & Mercer, 2003). We cannot assume, however, that students already have the skills and attitudes needed for dialogue that facilitates learning. Therefore, an instructional strategy for citizenship education in which dialogue is a central element - we will refer to this as dialogic citizenship education - should also support the development of these skills and attitudes. We consider the following skills and attitudes important for effective dialogue (Frijters et al., in press):

- *Exchanging*: being willing and able to express your own opinions and share these with others.
- *Co-constructing*: being willing and able to form your own opinions in a dialogue, utilizing the input of others, and contribute to the opinions of others.
- *Validating*: being willing and able to validate your own opinion and the opinion of others from the perspective of moral values.

#### *1.4 Group work versus whole-class instruction*

There are various ways of implementing dialogue in the classroom. A method often used to stimulate interaction is for students to work in small groups (Beane, 2002; Covell & Howe, 2001; Frijters et al., in press). When working in small groups, more students are able to participate in the dialogue than when the whole class is involved. In small groups the students also have more responsibility for the progress of the dialogue. They have to guide the dialogue themselves and solve possible conflicts. Whole-classroom discussion is also recommended (Tredway, 1995; Saye, 1998). Teachers guide the dialogue in whole-classroom discussion by asking questions. They can stimulate students to think about their opinions, and guide them towards a more profound, carefully reasoned opinion than in a dialogue without teacher guidance (Rojas-Drummond & Mercer, 2003).

#### *1.5 Research questions*

The study presented in this chapter investigated the effect of dialogic citizenship education on students' ability to take moral values and multiple perspectives into account when justifying their viewpoints. In this context we investigated the effect of the amount of group work. This study was on citizenship education as an integral part of history classes, so we were also interested in the effect of citizenship education on historical reasoning. The research focused on three questions:

- 1) Does dialogic citizenship education as an integral part of history classes enhance students' ability to take moral values and multiple perspectives into account when justifying their viewpoints?
- 2) Does the amount of group work in dialogic citizenship education contribute to students' ability to take moral values and multiple perspectives into account when justifying their viewpoints?
- 3) Is the integration of dialogic citizenship education into history classes to the detriment of students' ability to reason historically?

## 2. METHODS

### *2.1 Instructional materials*

In co-operation with a history teacher educator we designed two curriculum units of thirteen 45-minute lessons for students in the 8<sup>th</sup> grade of pre-university education. We tested both units in a pilot study. Two experienced teachers, with two classes each, participated in the pilot study. We observed the lessons and the teachers gave feedback after each lesson. The students were also interviewed. These data led to further refinements in the two curriculum units. (For a more detailed description of the two curriculum units, see Schuitema, Van Riessen, Veugelers & Ten Dam, submitted.) The teaching materials supplemented an existing history textbook that is widely used in the Netherlands. These extra materials consisted of a teachers' man-

ual and workbooks for the students. The main topic was the history of the United States of America from the first colonists until the beginning of the nineteenth century. Central issues were the founding of the USA, the position of the Native Americans, immigration to the USA, slavery and the Civil War.

*Recognizing and understanding moral values*

Both curriculum units paid systematic attention to moral values, which are defined as 'opinions, wishes or ideals on how people should behave towards each other'. Students learned to recognize and identify the moral values found in the learning materials. They studied, for instance, the text of the American Declaration of Independence from 1776 and parts of the 1788 Constitution, and were asked to indicate what values were incorporated in these texts. They also investigated what values they themselves and their fellow students considered to be important.

*Investigating multiple perspectives*

In both curriculum units students investigated multiple perspectives on moral issues. They were provided with several source materials reflecting different perspectives on a historical event or situation and also worked on assignments requiring them to empathize with a particular perspective. Some of the students studied, for example, source materials reflecting the perspectives of the Native Americans, while others studied sources reflecting the settlers' perspectives. They then discussed a number of statements and were asked to consider the views of a Native American and of a settler on these statements.

*Instructions for dialogue*

The curriculum units paid systematic attention to the skills and attitudes that students need to participate in a dialogue - exchanging, co-constructing and validating. From the very first lesson, they were encouraged to exchange opinions with other students, for example, by doing exercises in which they wrote down each other's opinions without the need for immediate agreement. Gradually we added activities aimed at co-construction and validation. When the students were co-constructing, they had to try to reach agreement and state which points they agreed and disagreed on. Then they had to determine the important values that had been involved in forming their opinions (validating).

*Differences between the two curriculum units*

The dialogue took place in small groups of students without a teacher in one curriculum unit. In the other unit, the dialogue generally occurred in a whole-classroom discussion under the teacher's guidance. The teachers' manual prescribed the time to be spent on each type of activity. In the unit aimed at group work, 65% of the available time was intended for working in small groups of 3-5 students or in pairs, and 15 % of the time was intended for classroom discussion. Twenty per cent of the time was intended for group work and 40% for classroom discussion in the other unit. The teachers in the group-work unit were instructed to give as little help as possible with the subject content and explicitly to guide the process of collaboration. Teachers were instructed to help the group as a whole when students asked for assistance.

In the classroom-discussion unit, the teachers had been instructed to guide the discussions by asking questions. They were also asked to ensure that as many students as possible participated in the dialogue (exchanging), interacted with each other (co-constructing) and underpinned their opinions (validating).

## 2.2 Design and procedure

We implemented a quasi-experimental research design, which compared two experimental conditions and a control condition. In the first experimental condition the teachers used the learning materials that emphasized working in small groups. (This condition is referred to below as the group-work condition). The teachers in the second experimental condition used the learning materials that emphasized whole-class discussion (the whole-class condition). In the control condition, teachers used the same history textbook as the teachers in the experimental conditions in a minimum of ten lessons (average 10.6).

We asked history teachers in schools in urban areas of the Netherlands to participate in the research project. Nine teachers from eight schools (fourteen classes in total) were willing to work with the additional teaching materials. The teachers were randomly assigned to one of the two experimental conditions. Five teachers from three schools (five classes) participated in the control condition.

Table 1. Number of classes per condition per school track

	High-track	Mid-track	Low-track	Total
Group-work	0	2	6	8
Whole-class	2	3	1	6
Control	2	2	1	5
Total	4	7	8	19

All the students were from the 8<sup>th</sup> grade of pre-university education (age 13-14). In the strongly streamed educational system in the Netherlands, the pre-university track caters for the cognitively more advanced students. It is usually divided into three sub-tracks, which we will refer to as the high-track, the mid-track and the low-track of pre-university education.<sup>2</sup> Students are selected for these tracks on the basis of their previous performance at primary and secondary school. Table 1 shows the dis-

<sup>2</sup> The high-track and the mid-track of pre-university education in the Dutch educational system are called *Gymnasium* and *Atheneum*. These two tracks are similar except that the high-track includes Latin and Greek. The low-track of pre-university education is a transition class. At the end of grade 8 some of the students in this track continue following pre-university education (the mid-track) and others transfer to general secondary education. The three tracks have the same history curriculum.

tribution of the classes from the three school tracks over the three conditions. There is no high-track in the group-work condition.

A total of 482 students participated in the research project. Of these students 49% was female, with a random distribution over the conditions ( $\chi^2 = 0.876$ ,  $df = 2$ ,  $p = .65$ ). Fifteen per cent of the students considered their ethnic identity to be non-Dutch or mixed, with a random distribution over the conditions ( $\chi^2 = 4.95$ ,  $df = 2$ ,  $p = .08$ ). The criterion adopted for the social economic status (SES) was the highest level of education of the parents. Twenty-one per cent of the students had a low SES (highest level of education was pre-vocational secondary education). Seven per cent had an average SES (senior secondary vocational education) and 72% had a high SES (higher education). These were distributed randomly over the conditions ( $\chi^2 = 5.263$ ,  $df = 4$ ,  $p = .26$ ).

A 45-minute session of pre-tests was held in all the conditions before the first of the thirteen lessons. Post-tests were held after the lessons. The students wrote an essay, in which they had to give their opinions on a statement. The class was randomly divided into two parts. One half of the students were given a topic directly linked to the teaching material, the other half a new topic that had not been discussed in the curriculum unit (transfer assignment, see section 2.3 'measurements'). Seven per cent of the students did not participate in the pre-tests; these students were randomly spread over the conditions ( $\chi^2 = 0.163$ ,  $df = 2$ ,  $p = .92$ ). Nine per cent did not participate in the post-measurements. This drop-out ratio was not evenly spread over the conditions ( $\chi^2 = 10.258$ ,  $df = 2$ ,  $p = .01$ ). In the group-work condition more students did not participate in the post-tests (14%) than in the whole-class condition (7%) and the control condition (4%). Table 2 shows an overview of the number of students that participated in the pre-tests and post-tests.

Table 2. Number of students participating in the pre-tests and post-tests per condition

	Pre-tests	Post-tests	
		Subject-matter related	Transfer
Group-work	214	85	112
Whole-class	131	57	73
Control	108	45	65
Total	453	187	250

### *Teacher training*

All the teachers in the experimental conditions received training in advance. The training consisted of an explanation of how to work with the teachers' manual and condition-specific instructions on the role of the teacher.

### *Implementation*

To check the implementation of the curriculum units, teachers kept a daily log. They recorded per activity whether the activity had been carried out as prescribed in the teachers' manual, and how much time had been spent on it. We used this data to calculate the percentage of time that the students spent working in small groups and on whole-class dialogue. In the group-work condition the average percentage of group work was 59 (varying from 55 to 64), and in the whole-class condition 24 (varying from 19 to 31) ( $t = 11.471, p < .001$ ). Then the percentages of time spent on classroom discussion were compared in both conditions. In the group-work condition, the teachers indicated that they had spent an average of 15% (varying from 11 to 18%) of the time on classroom discussion. In the whole-class condition this was 31% (29 to 34%) ( $t = 9.492, p < .001$ ). We also checked how many of the activities in the curriculum unit were carried out as prescribed in the teachers' manual. In the whole-class condition these percentages varied from 77 to 90%, with an average of 79%. In the group-work condition the percentages were slightly higher. The average was 87%, varying from 77 to 92%. A  $t$ -test did not show significant differences between the two conditions. On the basis of these results, we concluded that the curriculum units had been implemented accurately and that the two conditions differed markedly in the amount of group work and whole-class discussions.

## 2.3 Measurements

### *Dependent variables: essay assignments*

The effect of the two curriculum units on the students' ability to take moral values and multiple perspectives into account was investigated with the help of two different essay assignments. Both assignments consisted of a short introduction on a moral dilemma and a statement. In the first assignment (subject-matter related), the students had to consider the interests of the Native Americans as opposed to the interests of the immigrants from Europe and other continents to America. The statement was: 'To protect the way of life of the Native Americans, the American government should have controlled the arrival of new immigrants much sooner'. During the lessons the students had access to in-depth information on this issue and were able to use this information for the essay. The second essay assignment was intended to ascertain whether the students were able to apply what they had learned to a new subject. This essay assignment (transfer assignment) was therefore about a topic that had not been discussed in the lessons. The statement was: 'School uniforms back in the classroom!' The idea was to introduce school uniforms in the classroom to avoid students being bullied about the way they dress. Students in every class were randomly assigned one of the two essays assignments. The students were given ten minutes to discuss the statement in groups of four. Then they wrote down their opin-

ions on the statement individually. They were told beforehand that the score for the assignment would not be on the opinion itself but for their justification of it. A team of independent raters assessed the students' essays on *moral values and multiple perspectives*. The subject-matter related essays were also assessed for historical reasoning.

For the first aspect, *moral values*, the essay score was based on the number of arguments which referred to moral values, including the extent to which the students explicitly referred to a moral value; the more clearly a student referred to a general value transcending the specific context, the higher the score. An essay on school uniforms in which a student states, for example, that everyone should be able to make a personal decision about what to wear scored higher than an essay in which a student says that he wants to be able to choose for himself. An essay linking choosing your own clothes to freedom of expression scored even higher.

The aspect of *multiple perspectives* concerns the extent to which students discuss various perspectives in their essays. Attention was paid in the scoring not only to the number of perspectives but also to the degree of elaboration on the perspectives. The number of arguments and sub-arguments for each perspective was checked. Arguments for and against the statement often corresponded to two perspectives, but not always. When a student indicated, for example, that the drawback of a school uniform for him was that he would not be able to show off his new clothes, and that the advantage would be not to have to decide what to wear in the morning, these two viewpoints are really from the same perspective.

The aspect *historical reasoning* was assessed on the extent to which students took the context of the period in question into account (see e.g. Lee & Ashby, 2001). Did students in the subject-matter related essays apply the historical knowledge they had learned in the curriculum unit? Were students aware of the historicity of values? Did they use their historical knowledge in their argumentation and conclusions?

Each essay was scored separately for each aspect as a whole, using a method derived from the *comparison method* (Blok, 1986). The raters used *anchors*. These are model essays which the raters compared the other essays with. We chose (per aspect) three model essays: a weak, an average and a good essay. The three anchors had fixed scores - 50, 100 and 150 respectively. We then asked the raters to compare each essay with the anchors and score them on a scale ranging from 0 to 200. When a rater judged an essay to be better, for instance, than the average model essay, but not up to the standard of the good model essay, then this essay would score between 100 and 150 points.

To estimate the reliability of the essay scores we worked with the 'snake method' (Van den Bergh & Eiting, 1989). The essays were randomly divided into as many samples as there were raters. Each rater scored two samples. Rater 1 scored sample 1 and 2, rater 2 scored sample 2 and 3, and so on. The last rater scored the essays in the first and last sample. In this way, each essay was assessed by two raters, making it possible to estimate the reliability of each rater based on all the other raters. Different teams of raters were formed for each aspect to prevent one aspect influencing another. The teams comprised five raters for values and multiple perspectives and three raters, who were historians, for historical reasoning. The raters followed two to three-hour training sessions, which explained the particular aspect

and related criteria. The reliabilities of the essay assessments were estimated using a LISREL model (Jöreskog & Sörbom, 2002). The scores for the essays were calculated by taking the average of the two raters who had assessed the essay. The reliability of the average scores varied from 0.76 to 0.93 (see appendix A), indicating that the ratings were sufficiently reliable.

#### *Control variables*

Several pre-tests were conducted for those variables which were expected to correlate with the post-test. These pre-tests can be used for identifying possible relevant differences between conditions at the start of the curriculum unit.

- *Reasoning skills.* It can be expected that reasoning skills play an important part in the ability to formulate an opinion. To estimate the level of students' reasoning skills, three relevant scales (21 items) of the Cross-Curricular Skills Test were used (Meijer, Elshout-Mohr & Van Hout-Wolters, 2001): forming opinions, notions and beliefs and distinguishing facts and opinions. Cronbach's alpha over the three scales was .63.
- *Academic aptitude.* The scores from the Primary Education Final Test were collected from the schools. This is a standard test in the Netherlands, which children take in grade 6, the last year of primary education.
- *Attitudes towards dialogue.* Dialogue is an important educational strategy in both curriculum units. Students' attitudes towards dialogic learning at the start of the curriculum unit may influence the dependent variables. We used the 'Attitudes to dialogic learning scale' (Frijters et al., in press) to estimate students' attitudes on the exchange, co-construction and validation of opinions. ( $\alpha = .82$ )
- *Concern for others.* Forming an opinion while considering other people's opinions may be influenced by the extent of the commitment to and concern for other people. To measure the differences in social commitment we used a scale of the Child Development Project (Concern for Others Scale) (Solomon et al., 2001). ( $\alpha = .73$ ).
- *School culture.* The way in which teachers and students behave towards each other in school may also influence students' opinion forming. School culture was estimated with the 'School Culture Scale' (Higgins-D'Allesandro & Sath, 1997) ( $\alpha = .80$ ).
- *Classroom climate.* A final aspect that may influence students' opinions is the climate in the classroom when social and political matters are discussed. We used the IEA Measure of classroom climate for discussion (Torney-Purta, Lehmann, Oswald & Schulz, 2001). This scale measures whether students are encouraged to express their opinions in the classroom and feel free to do so ( $\alpha = .69$ ).

#### 2.4 Analyses

Students worked on the essay assignments in small groups. They first discussed the statements in their groups and exchanged arguments. We analysed the essay scores with the aid of multi-level regression analysis (MLwiN 2,02: Rasbash, Browne,

Cameron & Charlton, 2005) because of the hierarchically nested structure of the data. Models with three levels were compared (student, group, and class). The assessment of the essays resulted in five different scores: *values*, *multiple perspectives*, and *historical reasoning* in the subject-matter related assignment, and *values* and *multiple perspectives* in the transfer assignment. The effect of the condition was investigated in a multivariate, multi-level analysis with these five scores as dependent variables. The variances for the different dependent variables were taken together

As the average scores for the essays differed for each of the three tracks of pre-university education, and these tracks were unevenly distributed over the conditions, we investigated the effect of the condition within the three school tracks. In the first model the condition was included with the school track. To control for individual differences in gender, SES, and ethnic identity, these student characteristics were added in a step-by-step procedure in which only significant variables were included in the model (model 2). For model 3 we checked in the same way whether the control variables could contribute to improving the model and again only significant variables were included.

Each variable was first added to the model with separate regression weights for each dependent variable. We then checked whether these regression weights differed significantly for each dependent variable. When they did not differ significantly, we can assume that the effect of the variable concerned is the same for every dependent variable. In that case, to keep the model as efficient as possible, the regression weights were replaced by one regression weight for all the dependent variables together.

### 3. RESULTS

In this section we present the results of the multi-level analyses: to what extent does dialogic citizenship education enhance students' ability to take moral values and multiple perspectives into account when justifying their viewpoints, and to what extent is this process influenced by the amount of group work? The ability to take values and multiple perspectives into account was operationalized as students' references to moral values in the essays and the discussion of multiple perspectives in the essays. Another research question answered in this section is to what extent citizenship education in history classes affects students' ability to reason from a historical point of view. Table 3 presents the essay scores. In model 1 only the regression weights for condition and school track have been included. The average score of low-track students in the control condition is the intercept, and the group-work and whole-class conditions are indicated as deviations. The average score on the aspect of *values* in the subject-matter related assignment is estimated as 54.2 for low-track students. For this assignment the low-track students in the group-work condition had higher scores for referring to values than the low-track students in the control condition. (The difference is 23.4 points.) The students in the whole-class condition referred less to values in their essays than the students in the control condition (- 9.8), but this difference is not significant. The difference between the group-work

Table 3. Results of the multi-level analyses of the essay assignments

			Model 1		Model 2		Model 3	
			Mean	SE	Mean	SE	Mean	SE
Subject-matter related								
Values	low-track	control	<b>54.2</b>	(10.9)	<b>52.5</b>	(10.7)	<b>52.3</b>	(10.6)
		group-work <sup>1</sup>	<b>23.4</b>	(11.6)	<b>23.7</b>	(11.2)	<b>24.2</b>	(11.1)
		whole-class <sup>1</sup>	-9.8	(12.5)	-9.5	(12.1)	-9.0	(11.9)
		gender (girl) <sup>2</sup>			3.8	(4.7)	3.7	(4.6)
Multiple perspectives	low-track	dialogue					7.6	(5.2)
		control	<b>56.4</b>	(11.0)	<b>50.9</b>	(10.9)	<b>50.7</b>	(10.8)
		group-work <sup>1</sup>	<b>25.6</b>	(11.9)	<b>26.1</b>	(11.5)	<b>26.8</b>	(11.3)
		whole-class <sup>1</sup>	-14.3	(12.7)	-13.9	(12.3)	-13.4	(12.1)
Historical reasoning	low-track	gender (girl) <sup>2</sup>			<b>11.3</b>	(5.0)	<b>11.2</b>	(5.0)
		dialogue					10.6	(5.6)
		control	<b>61.9</b>	(10.9)	<b>62.1</b>	(10.7)	<b>62.2</b>	(10.6)
		group-work <sup>1</sup>	6.2	(11.6)	6.4	(11.2)	6.3	(11.1)
		whole-class <sup>1</sup>	<b>-25.1</b>	(12.5)	<b>-24.9</b>	(12.1)	<b>-24.4</b>	(11.9)
Transfer Values	low-track	gender (girl) <sup>2</sup>			-0.1	(4.7)	-0.2	(4.7)
		dialogue					-1.0	(5.2)
		control	<b>70.9</b>	(10.9)	<b>66.5</b>	(10.7)	<b>67.1</b>	(10.5)
		group-work <sup>1</sup>	15.6	(11.6)	15.3	(11.2)	16.2	(11.0)
Multiple perspectives	low-track	whole-class <sup>1</sup>	2.8	(12.5)	2.2	(12.1)	2.0	(11.9)
		gender (girl) <sup>2</sup>			<b>10.2</b>	(4.6)	8.6	(4.6)
		dialogue					<b>14.2</b>	(5.8)
		control	<b>58.2</b>	(10.6)	<b>51.2</b>	(10.3)	<b>51.2</b>	(10.2)
		group-work <sup>1</sup>	<b>40.1</b>	(11.2)	<b>39.5</b>	(10.8)	<b>39.3</b>	(10.6)
Variance	low-track	whole-class <sup>1</sup>	6.9	(12.1)	5.8	(11.6)	6.3	(11.5)
		gender (girl) <sup>2</sup>			<b>16.2</b>	(4.0)	<b>11.2</b>	(4.0)
		dialogue					-0.53	(4.9)
		mid-track <sup>3</sup>	<b>25.8</b>	(8.4)	<b>24.7</b>	(8.1)	<b>24.7</b>	(8.0)
Variance	low-track	high-track control <sup>3</sup>	21.35	(14.5)	22.1	(13.9)	22.0	(13.7)
		high-track whole-class <sup>3</sup>	<b>68.6</b>	(14.1)	<b>68.7</b>	(13.6)	<b>67.2</b>	(13.4)
		level 3 (class)	<b>165.6</b>	(69.1)	<b>150.0</b>	(63.6)	<b>143.7</b>	(61.8)
Variance	low-track	level 2 (group)	<b>143.5</b>	(42.3)	<b>139.8</b>	(41.1)	<b>144.4</b>	(41.4)
		level 1 (student)						
		subject-matter related values	<b>665.4</b>	(79.6)	<b>663.9</b>	(79.3)	<b>655.2</b>	(78.4)
		multiple p. hist. reasoning	<b>817.5</b>	(96.1)	<b>800.1</b>	(94.1)	<b>784.4</b>	(92.4)
		multiple p. hist. reasoning	<b>663.5</b>	(79.5)	<b>662.4</b>	(79.3)	<b>656.7</b>	(78.7)
		transfer values	<b>1068.1</b>	(107.2)	<b>1045.1</b>	(104.7)	<b>1017.9</b>	(102.2)
Fit	low-track	multiple p.	<b>768.2</b>	(79.4)	<b>720.5</b>	(74.6)	<b>717.4</b>	(74.3)
		improvement	9494.0		9471.3		9459.2	
Fit	low-track	difference in degrees of freedom	<b>59.6</b>		<b>22.7</b>		<b>12.1</b>	
		p-value	13		5		5	
		p-value	p < .001		p < .001		p = .034	

**Bold type** is significant ( $p < 0.05$ )

<sup>1</sup> Difference with control group

<sup>2</sup> Difference between boys (= 0) and girls (= 1)

<sup>3</sup> Difference with low-track

condition and the whole-class condition can be calculated with the data from the table. It is 33.2 and is significant ( $\chi^2 = 9.2$ ,  $df = 1$ ,  $p = .002$ ).

Table 3 also shows that the low-track students in the group-work condition took more perspectives into account than the low-track students in the control condition. The students in the group-work condition had significantly higher scores for *multiple perspectives*, both on the subject-matter related assignment and on the transfer assignment. The low-track students in the group-work condition scored also higher than those in the whole-class condition. The difference in multiple perspectives in the subject-matter related assignment is 39.88 ( $\chi^2 = 12.79$ ,  $df = 1$ ,  $p < .001$ ). The difference in multiple perspectives in the transfer assignment is 35.13 ( $\chi^2 = 8.78$ ,  $df = 1$ ,  $p = .002$ ) The average scores for values in the transfer assignment do not differ significantly between the three conditions.

Concerning historical reasoning, Table 3 shows that the low-track students in the group-work condition performed as well as the low-track students in the control condition. The students in the whole-class condition, however, scored significantly lower than the students in the control condition (- 25.1). The difference between the group-work condition and the whole-class condition must also be calculated separately here. The difference is 31.3 and is significant. ( $\chi^2 = 8.20$ ,  $df = 1$ ,  $p = .004$ ).

The effect of the school track did not appear to differ over the various dependent variables. We therefore included one regression weight for all the dependent variables for the different school tracks. For the mid-track, only a main effect was significant and included in the model, which makes the picture for mid-track students the same as for low-track students. All of the averages for mid-track students in the model are 25.8 points higher than for low-track students. Since there were no high-track students in the group-work condition we do not have a main effect for these students. The effect for the high-track has, therefore, been split into two, and separate regression weights have been added for the students in the whole-class condition and the control group. In Table 3 we see that high-track students in the whole-class condition scored significantly higher on all the dependent variables (68.6) than the low-track students in the same condition. In the control condition, however, high-track students did not score higher than the low-track students. In the second model, gender was added. The regression weights for gender differed significantly per dependent variable and could not be taken together. Ethnic identity and SES did not appear to be significant and were removed from the model. Table 3 shows that girls scored significantly higher for multiple perspectives in both essay assignments. Girls also referred more often to values than boys in the transfer assignment. The effects of condition and school track remained the same.

In model 3 the pre-test on attitudes towards dialogic learning was added. Here too the regression weights for the various dependent variables could not be taken together. The remaining control variables did not appear to be significant. The effect of attitudes towards dialogic learning only appeared to be significant to the aspect of values in the transfer assignment. After the addition of this control variable the significant effect of gender disappeared in the transfer assignment. Adding this variable did not change the remaining effects.

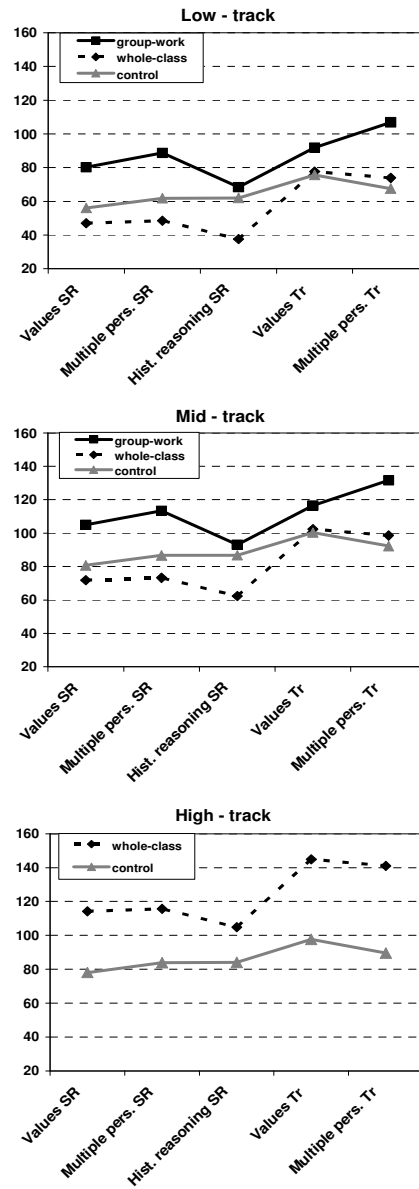


Figure 1. Estimated averages of the essay scores for girls in the low, mid, and high track of pre-university education.

Figure 1 shows the averages for the different conditions of the five dependent variables per school track as estimated in model 3. It only shows the mean scores for the girls. The pattern is the same for boys, but then both scores for the aspect of multiple perspectives are 11.2 lower. The figure shows that the pattern for low-track students is the same as for mid-track students, but the scores for mid-track students are higher. The students in the group-work condition scored higher than the students in the whole-class condition and in the control condition for all the variables. However, only those differences in the aspect of values in the subject-matter related assignment and in the aspect of multiple perspectives in both the subject-matter related assignment and the transfer assignment were significant. The students in the whole-class condition scored the same as the students in the control condition, except for historical reasoning; the students in the whole-class condition had lower scores for this.

For the high-track students we see a different pattern. As mentioned before, there was no group-work condition for high-track students. The average scores of the high-track students in the whole-class condition on all the dependent variables are higher than those of the high-track students in the control condition. Even the smallest difference between these two groups on historical reasoning is significant ( $\chi^2 = 5.15$ ,  $df = 1$ ,  $p = 0.023$ ).

#### 4. DISCUSSION

In this study we investigated the effect of dialogic citizenship education as an integral part of history classes on students' ability to take moral values and multiple perspectives into account when justifying their viewpoints. We also investigated in particular the effect of the amount of group work. We designed two curriculum units aimed at stimulating students to form an opinion, taking into account conflicting values and different perspectives. The units differed in the balance between group work and classroom discussion. The results firstly show that it is worthwhile to pay systematic attention to moral values and multiple perspectives in school subjects. History education offers good opportunities for modelling citizenship education. The effectiveness of the two different instructional approaches to citizenship education, however, differs. Students who do relatively a lot of work in small groups refer to values in their essays more often and more explicitly, and are better able to validate the different perspectives, in comparison to students who have done more work in whole-class situations.

The effect of dialogic citizenship education differs between the three tracks of pre-university education. In the low-track and mid-track classes the amount of group work plays an important role in enhancing students' ability to take values and multiple perspectives into account. The low-track and mid-track students in the whole-class condition did not have higher scores than those in the control condition. In contrast, the students in the group-work condition had higher scores than the students in the two other conditions for a number of aspects. These students scored higher for taking multiple perspectives into account. This applies to both the subject-matter related assignment and the transfer assignment. These results show that when

students work in small groups relatively often they are more able to include different perspectives when justifying their opinions than students who work in small groups less often. For the aspect of *values* a distinction must be made for the low-track and mid-track students between the subject-matter related assignment and the transfer assignment. It was only in the subject-matter related assignment that the students in the group-work condition scored higher on values than the students in the control and whole-class conditions.

Regarding high-track students, it appears that students in the whole-class condition had higher scores for all the aspects of the essays rated than students in the control condition. Since there were no high-track students in the group-work condition we can only make a hypothetical prediction of the effect of group work on these students. As the students in the group-work condition scored considerably higher on a number of aspects, we can assume that a lot of group work would not be disadvantageous for these students.

The question whether integration of citizenship education into history classes is to the detriment of students' ability to reason historically cannot be answered in a general sense. This depends on the teaching method used. The scores for historical reasoning of the students in the low-track and mid-track classes who did relatively a lot of work in small groups appeared to be as high as the scores of the students from the control condition. However, the students in the whole-class condition had lower scores for historical reasoning than the students from the group-work condition and the control condition.

The effects of dialogic citizenship education on the ability of the low-track and mid-track students to take moral values and multiple perspectives into account were only evident in the group-work condition. This may be partly due to the form of the essay assignment. Before writing down their individual opinions, the students had ten minutes to discuss the statement in groups of four. The students in the group-work condition were more used to working in small groups than the students in the whole-class condition. Hence the results may at least be partly due to the experience gained by learning to work together in groups in the curriculum units rather than to the specific teaching methods which paid attention to the skills and attitudes students need to formulate arguments. We therefore recommend that future research pays attention to the influence of the quality of the group dialogues on students' ability to take values and multiple perspectives into account.

The fact that there was no effect on values in the transfer assignment for students in the low-track and mid-track classes is a result that we often see in educational research, but is nonetheless disappointing. The students who had worked relatively often in small groups referred more often to values, when they had been taught the topic in question in the lessons and were familiar with it. When they were confronted with new subject matter, however, they did not make more references to moral values than students in the control group who had not participated in citizenship-education lessons. Many students were not able to decontextualize a dilemma, such as wearing a school uniform, from the specific context of their own everyday environment and to consider it in a broader social context. In other words, their ability to refer to values when forming an opinion remains directly linked to the subject that has been taught. No transfer occurred. We therefore recommend that future re-

search pays explicit attention to integrating transfer-inducing qualities of instructional-learning processes into a group-work-oriented approach to dialogic citizenship education. (See e.g. Elshout-Mohr, Van Hout-Wolters & Broekkamp, 1999; Volman & Ten Dam, 2000.)

Another explanation for the non-transfer of the ability to refer to values is a matter of moral sensitivity (Tirri & Pehkonen, 2002). Did the students realize that the transfer assignment concerned a moral dilemma? Whereas the curriculum units paid attention to values through exercises which explicitly stated that the aim of the assignment was to recognize moral values, this was not clear in the instructions for the final-essay assignments. The express purpose of the assignment was that students would refer to moral values even in subjects that were new to them. For many students, this may have been a bridge too far.