The Voice of Holland: Allograph Production in Written Dutch Past Tense Inflection

de Bree, E.; van der Ven, S.; van der Maas, H.

DOI
10.1080/15475441.2016.1217777

Publication date
2017

Document Version
Final published version

Published in
Language Learning and Development

LANGUAGE LEARNING AND DEVELOPMENT
2017, VOL. 13, NO. 3, 215–240
http://dx.doi.org/10.1080/15475441.2016.1217777

The Voice of Holland: Allograph Production in Written Dutch Past Tense Inflection

Elise de Bree[a], Sanne van der Ven[b,c], and Han van der Maas[c]

[a]Research Institute of Child Development and Education, University of Amsterdam; [b]Department of Child, Family and Education Studies, Utrecht University; [c]Department of Psychological Research Methods, University of Amsterdam

ABSTRACT

According to the Integration of Multiple Patterns hypothesis (IMP; Treiman & Kessler, 2014), the spelling difficulty of a word is affected by the number of cues converging on the correct answer. We tested this hypothesis in children's regular past tense formation in Dutch. Past tenses are formed by adding either-*de* or-*te* to a verb stem. Despite instruction, children often choose the wrong allograph. In a large dataset (227 items, together completed 392,802 times) from an online language program we assessed whether morphophonological and orthographic cues determine differences in difficulty and explain error patterns. Regression analyses established that inflection difficulty was affected by number of converging cues, especially morphophonological and orthographic cues. Error analyses further showed that allograph errors were prominent when graphotactic frequency and especially voicing probability collided with the correct answer. The results match and specify the IMP. Proposals are made to use this knowledge in educational practice.

Becoming fully literate entails being able to read as well as spell properly. Learning to spell correctly is a protracted process in many languages. It demands integration of phonological, morphological, and orthographical information. Theories of spelling have proposed different interpretations of when and how the different components of knowledge are integrated in spelling (e.g., Nunes, Bryant, & Bindman, 1997; Pacton & Deacon, 2008; Templeton & Morris, 2000; Treiman & Cassar, 1996; 1997; see Treiman & Kessler, 2014; for an overview). One viewpoint is that these skills are used and integrated gradually from an early age onwards (e.g. Kemp & Bryant, 2003; Pacton & Deacon, 2008; Pollo, Treiman, & Kessler, 2008; Treiman & Kessler, 2006). Within this viewpoint, the Integration of Multiple Patterns (IMP) model, a recent model by Treiman and Kessler (2014), holds that children acquire spelling patterns on the basis of general all-purpose mechanisms, such as statistical learning, but also linguistic knowledge, as well as instruction. Treiman and Kessler propose that "Children tend to do well when several patterns converge on the correct answer. They have more difficulty when this isn't the case" (p. 98). The present study is an in-depth evaluation of the IMP by focusing on the acquisition of written Dutch past tense, in which converging and colliding cues are present.

Dutch orthography is relatively transparent, characterized by a substantial degree of phoneme-grapheme correspondences. However, the transparency for reading is higher than for spelling (Bosman, Vonk, & van Zwam, 2006), which is one of the reasons why spelling acquisition is slower than reading. This asymmetry warrants the study of orthographical and morphological rules next to phonological knowledge for spelling acquisition. A particularly interesting case in this respect is the

CONTACT Elise de Bree e.h.debree@uva.nl Research Institute of Child Development and Education, University of Amsterdam, PO Box 15780, 1001 NG Amsterdam, The Netherlands.

The first two authors contributed equally to the study.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hlld.

spelling of regular verb inflections. A recent literature review has established that verb spelling is the largest stumbling block in Dutch spelling education (see Bonset & Hoogeveen, 2009).

Past tense spelling is an area of difficulty within verb spelling for Dutch children (Assink, 1987; Neijt & Schreuder, 2007) and even adults (DeSchryver, Neijt, Ghesquière, & Ernestus, 2013; Ernestus & Baayen, 2001). However, not all past tense verbs spellings are equally difficult. For some verbs, phonological, morphological, and orthographic cues converge on the correct spelling (consistent verbs), whereas in others (semi-consistent and inconsistent verbs), cues collide. These verb types and the (dis)agreement between the cues are presented in Table 1 and will be discussed below. The IMP predicts that items with colliding cues would be more difficult to master than those with cues that converge on the same and correct answer. In this study we address this matter by investigating which factors contribute to the difficulty of Dutch regular past tense formation and which error patterns are attested for the different types of verbs. Not only does this allow us to evaluate the IMP, it also allows us to identify elements that could be incorporated in education.

Similar to English, regular past tenses in Dutch are created by the adding a suffix to a verb stem. This suffix can take two different forms (called allomorphs in spoken form; allographs in written form): Verb stems ending in an underlyingly voiceless obstruent take unvoiced *-te* (singular) or *-ten* (plural), all other cases take voiced-*de(n)* (Booij, 1995; Trommelen & Zonneveld, 1979). For example, the infinitive *scheppen* (to dig) contains a stem ending in a voiceless obstruent (*schep*) and subsequently takes-*te(n)* (*schepte(n), dug*). In contrast, *rennen* (to run) with a stem-final (voiced) nasal (*ren*), takes *-de(n)* in the past tense (*rende(n)*). Acquisition of regular past tenses in Dutch thus demands knowing when to inflect and spell a verb with-*de(n)* and when with-*te(n)*. To facilitate reading the present paper, we now refer to-*de(n)* as-*de* and to-*te(n)* as-*te*.

The spelling rule on adding the voiceless and voiced allographs is taught in the higher primary school years. It states that for past tense inflection, it should be checked whether the letter preceding the-*en* in the infinitive is part of the consonants in the mnemonic "'t kofschip". If so, then the past tense is formed by adding-*te* (*schepte*); otherwise it is-*de* (*rende*). Although this rule explicitly addresses morphology, phonology, and spelling, its acquisition is protracted in (spoken and) written language (DeSchryver et al., 2013; Neijt & Schreuder, 2007). The difficulty resides in being able to inflect for past tense (morphology) but especially in selecting the correct allograph (involving morphophonology and orthography). Part of the difficulty of this selection is due to the fact that Dutch is characterised by final devoicing: spoken word-final obstruents (i.e., plosives and fricatives) are never voiced: *bed* (bed) is pronounced /bƐt/. Similarly, verbs with stems ending in plosives and fricatives are affected by devoicing. A verb such as *tobben* (to worry) has stem *tob*, pronounced /tɔp/, but past tense *tobde* (/tɔbdə/. This inconsistency affects the ease with which the correct allograph can be detected in past tense production (i.e., production of **topte* instead of *tobde*).

Table 1. Categories of past tense consistency and the cues converging with correct past tense allograph.

		Verb type Consistent (Nasals, glides, vowels) n, m, l, r, j, i *-de*	Semi-consistent (Plosives) p, t, k *-te*	b, d *-de*	Inconsistent (Fricatives) f, s, ch *-te*	v, z, g *-de*
Spoken	Infinitive	Yes	Yes	Yes	Yes	Yes
	Stem	Yes	Yes	No	Yes	No
	Voicing probability	Yes	Mixed	Mixed	Mixed	Mixed
Written	Infinitive	Yes	Yes	Yes	Yes	Yes
	Stem	Yes	Yes	Yes	Yes	Mixed
	Graphotactic frequency	Yes	Yes	Yes	Mixed	Mixed
Number of cues converging		6/6	5/6	4/6	4/6	2/6

Relatively little in-depth data has been reported on Dutch past tense allograph production in children. In a study by Neijt and Schreuder (2007), a-*de* preference was found. The target verb was presented visually to children in a booklet, such as a picture of a person drawing for the verb *tekenen* (to draw). The verb was also presented orthographically with dots in the middle: *teken . e*. Children were asked to circle the letter belonging to the verb ("t" or "d") at the top of the page. Neijt and Schreuder found that the "d" was often selected correctly in-*de* verbs (with 28% incorrect—te realizations), but was also often selected incorrectly (48%) in-*te* verbs. One explanation could be that hypercorrection takes place. Children apply devoicing in oral production, but are taught to write a "*d*", in nouns (*bed* "bed") and past participles (*gehoord* "heard"). They could thus automatically rely on -*de* production in past tenses too. This process of hypercorrection (mentioned in, e.g., Kerkhoff, de Bree, & Buesseler, 2014; Neijt & Schreuder, 2007) entails that a -*de* preference should be attested across the board, as children would infer a pattern on the basis of the spelling they see and would apply this strategy in all instances.

In contrast to hypercorrection, the IMP predicts that there is no mere preference for-*de*, but that this preference is dependent on spelling of specific verbs as well as phonological, morphological and orthographic patterns. The general pattern of devoicing in Dutch, for instance, leads to different degrees of past tense verb consistency. Verbs that are not affected by devoicing show full spoken and written consistency: these are the verbs with stems ending on glides, liquids, and nasals (see Appendix 1 for phoneme categories and consistency). It is therefore easy to deduce the voicing value and allograph: always-*de*. Verbs with stems ending in plosives (e.g., /b/) and fricatives (e.g., /v/) are affected by devoicing. This leads to ambiguity, as the *tobben* example showed. Devoicing is present in the spoken but not the written form of the verb. As a consequence, the stem provides the correct orthographic cue for the-*de* allograph, but the pronunciation of stem and past tense can be inconsistent with spelling. For verb stems ending in a fricative there is an even higher degree of inconsistency. Voiced fricatives take-*de* as a past tense allograph (*leven* live takes-*de* in past tense). However, devoicing of stem-final voiced fricatives is present in both the spoken and the written forms of the stem as well as in the written form of the past tense: [z] and [v] are spelled as "s" and "f", respectively (*leefde* for /levdə/). The pronunciation of the past tense of these voiced fricatives is thus inconsistent with the spelling. The stem is ambiguous: verb stems ending in a fricative are always voiceless, regardless of the voicing value of the infinitive, and knowledge of the infinitive is again necessary to spell the past tense inflection correctly: *eis* (demand) takes-*te* because the infinitive is written and pronounced with an unvoiced "s" *eisen*, but *reis* (travel) takes-*de* as the infinitive is *reizen* with a voiced "z". We expect this devoicing to lead to higher difficulty and a high degree of allograph errors for verbs affected most by devoicing, thus especially the voiced and voiceless fricatives (**reiste* and **eizde*), to a lesser degree the plosives, while we expect low difficulty for verbs ending in nasals, glides, and liquids. We refer to the latter three as the "consistent group", those with plosives are "semi-consistent" verbs and those ending in fricatives constitute the "inconsistent" verb group.

An even more refined cue influences allograph selection of the semi-consistent and inconsistent verb groups (i.e., the plosives and especially the fricatives): voicing probability in speech. This voicing probability (Ernestus & Baayen, 2001, 2003, 2004) is the degree to which words with a specific pattern of stem-final phonemes are underlyingly voiced. It is determined by vowel length, the type of phoneme preceding the final consonant and the type of final consonant. For instance, voicing probability is lowest (.135) for a verb stem ending in short vowel /Ɛ/(spelled e) and final plosive p, as in *schepte* (dug), and highest (.953) for a stem ending in short vowel /Ɛ/followed by fricative /χ/, as in *legde* (layed). Adults have been shown to use knowledge of voicing probability when asked to inflect ambiguous pseudoverbs: they were more likely to inflect pseudoverb stem *tep* (voicing probability .135) as *tepte*, and not as *tebde*, while they preferred to inflect pseudoverb *meg* (voicing probability .953) as *megde* and not as *megte* (Ernestus & Baayen, 2003). They thus use their implicit knowledge of voicing probability to resolve the ambiguity in past tense formation based on the stem. Although the influence of this type of statistics has been reported to affect real and pseudoword past tense realisation of adults (e.g., DeSchryver et al., 2013; Ernestus & Baayen, 2001, 2003, 2004), the extent of this influence on children's past tense production has not been documented. If voicing probability is found to affect past tense production, this means that children are sensitive to voicing value, rhyme structure, and voicing probability in general in their morphology and apply this analogy to verbs. This would be anticipated on the basis of IMP, as voicing probability would be one of the multiple statistical regularities that is taken into account in the acquisition of spelling. Furthermore, verbs in which this voicing probability cue collides with the correct answer should be harder to inflect correctly. For instance, a verb such as *bruisen* "sparkle" takes-*te* on the basis of the pronunciation and spelling of the infinitive. However, the voicing probability value is .755, favoring the voiced counterpart (**bruisde*).

An additionally previously unexplored cue that can be assumed to influence Dutch past tense spelling is the frequency of the stem-final grapheme and the allograph consonant, such as relative frequency of *fd* and *ft*. For the (semi-)consistent verbs, this measure of graphotactic frequency converges on the correct answer. For the inconsistent verbs, however, this frequency measure can collide with the correct answer. For instance, the combination of graphemes *fd* in past tense and participle verb spelling (in verbs such as *leefde*—lived, but also past participles, such as *geleefd*—past participle of live) is much more frequent that the combination of graphemes *ft* based on CELEX (Baayen, Piepenbrock, & van Rijn, 1993). This frequency of *fd* (85%) compared to *ft* (15%) might lead to errors with fricatives that should be spelled with take-*ft*, such as *besefte* realized (i.e., **besefde*). Thus, although there is no devoicing between infinitive *beseffen* and stem *besef* and although voicing probability of this verb is low (.081), favoring-*te*, the graphotactic frequency favors *fd*. Expectations for the IMP are twofold. First, in this hypothesis, orthographic information would be another source of information that is integrated in children's spellings. Second, items in which this orthographic cue converges with the correct answer will be easier to produce correctly than those in which it does not. Findings in this vein have been reported in French (Pacton, Fayol, & Perruchet, 2005) and English (Deacon, Leblanc, & Sabourin, 2011), showing that children are sensitive to multiple (i.e., morphological and orthographical) regularities. Deacon et al. (2011) further showed that when these cues collide, six-year-old children relied more on the orthographic cues, but an interaction between morphology and orthography was also attested.

Present study

In the present study, Dutch past tense spelling was used as an in-depth test case for the IMP. Although previous studies have found that different cues are involved in spelling outcomes (e.g., Deacon et al., 2011; Pacton et al., 2005; Treiman & Kessler, 2006), the IMP as such has not been tested. Here, it was assessed to what extent the number of converging cues contribute to spelling. Furthermore, the relative importance of each cue was investigated: phoneme category, devoicing, consistency between spoken and written form, voicing probability, and orthographic frequency (graphotactic frequency). Additionally, error analyses looked into the question of whether allograph errors were more frequent in verbs in which cues collide (inconsistent verbs) than those that were more consistent (semi-consistent and consistent verbs). Data were analyzed from the *Verb Game* in *Language Sea* (Taalzee, Oefenweb, 2014), a computer-adaptive system used by many Dutch children in schools to practice language. In the task, a child is asked to provide the correct tense of a verb, presented as an infinitive between parentheses in a sentence. The items are presented adaptively such that each child only receives items tailored to their ability. A rating system, explained in more detail in the Method section, assigns a difficulty value to each verb (n = 227).

Regression analyses were used to investigate which characteristics contribute to item difficulty. In these analyses we investigated whether the morphophonological characteristics of consonant type (consistent, plosive, or fricative), voicing probability, and graphotactic frequency explain variance in difficulty when controlling for other effects that are less directly related to the morphophonological nature of the verb. We included other item characteristics as control variables. These control variables consist of variables that potentially influence spelling of the past tense in general, as well as variables that are relevant to the task design. Variables that affect past tense spelling in general are word length, spelling difficulty of the infinitive, singular/plural spelling, and token frequency. Longer words are expected to be more difficult to spell correctly, as are verbs with more difficult infinitives (e.g. DeSchryver et al., 2013). Furthermore, plurals are probably more difficult to spell correctly than singular verbs, as they also demand syntactic agreement. The morphological distinction of singular and plural is visible in spelling e.g., *ik schepte* (I dug) and *wij schepten* (we dug). With respect to token frequency, it has been found that highly frequent verbs are easier to inflect than those of lower frequency, also in children's oral past tense production (e.g., Matthews & Theakston, 2006; Rispens & de Bree, 2014). Task-specific characteristics that might influence productions are capitalization, with targets demanding a capital (the first word of the sentence) being more difficult as they add another demand to the task and two words (e.g., *opgroeien—groeide op*; grow up—grew up) being more difficult to type correctly than one word past tenses (e.g., *schepte* dug). Furthermore, different time cues are presented, i.e., indicating past tense through "v.t." (*verleden tijd*, past tense) or implicitly in the text ("yesterday,. . ..)"). This mode of presentation might influence the past tense production, although the direction cannot be predicted beforehand. Importantly, the aim is to assess whether the variables under scrutiny here (stem-final consonant type, voicing probability, and graphotactic frequency) contribute to past tense spelling after controlling for the impact of these control variables. Assessment of the contribution of these different cues while controlling for other variables has not been conducted before.

Error analyses are included to test whether the nature of the errors followed a predictable pattern expected from the differences in difficulty. Wrong allographs are not the only error type possible, e.g., errors can also be morphological (e.g., present tense inflection rather than past tense), vowel change in an analogy to irregular verbs (e.g., *kneeg instead of *knaagde*, gnawed), or they can be typos unrelated to the past tense. We expect the degree of wrong allograph selection to be related to the type of verb: fewest for consistent verbs, more for plosives and most for fricatives. Wrong allograph selection is also anticipated to be related to the voicing probability of these verbs (especially high when voicing probability is inconsistent with the correct realisation) as well as to graphotactic frequency. The findings on the assessment of the role of phonology, morphology, and orthography in past tense spelling of this study can be used in education to tailor instruction to especially those cases when spelling rules conflict with the implicit (linguistic and orthographic) knowledge children receive on the basis of other cues.

Method

As this study is based on item-based analyses obtained from a computer-adaptive program, the structure of the Methods section is slightly unconventional. Information about the participants in *Language Sea* and the Verb Game is followed by information on the general interface (*Language Sea* Interface) and algorithms of *Language Sea* (*Language Sea* Computer Adaptive Technology). Finally, information is provided on the past tense items in the Verb Game, the verb-related variables of interest, and the planned analyses.

Language Sea participants

Language Sea participants are primarily children whose schools have bought accounts for all students/pupils, providing a fairly good representative sample of Dutch children. Participating families and schools (approximately 500) gave permission for the use of the data of their pupils for research purposes; the schools accepted the responsibility to inform the parents, who were given the opportunity to refuse the use of their children's data.

During the period of data collection (October 2012—March 2014), the Verb Game was played by 38,120 children who on average played 9.3 items each. For the purpose of the present article we only

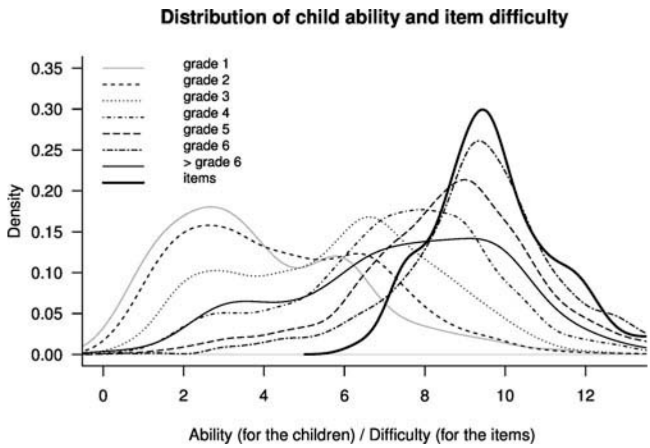

Figure 1. Distribution of the ability estimates of the children (by grade) and the difficulty estimates of the past tense items. Both are on the same scale. The plot is a density plot, which can be interpreted as a smoothed histogram.

looked at the past tense items: a subset of all items within the Verb game. Past tense items are only presented in this game once children have obtained high enough scores. Effectively, this means that the easiest past tense items will be presented in grades beyond kindergarten and the most difficult items in later grades. As an illustration, Figure 1 shows an overview of the distributions of both children's ability and past tense item difficulty (for an explanation of ability and difficulty see *Language Sea* Computer Adaptive Technology) obtained in February 2014: halfway through the school year. The horizontal axis displays the ability of the children and the difficulty of the items: the further towards the right, the higher the ability of the child and the difficulty of the item. The ability of the children is displayed by grade; the density plot can be interpreted as a smoothed histogram. The thick black line displays the difficulty of the items. The difficulty and ability values are based on Item Response Theory, meaning that the scale itself is arbitrary, but the difference between difficulty and ability indicates the probability of success. When ability and difficulty are equal, the child has a .50 chance of answering the item correctly.

As explained further in the next section (see *Language Sea* Computer Adaptive Technology), children are presented with items that are approximately 1–2 points lower than their own ability. Figure 1 confirms that the past tense items are rather difficult for children: the majority of past tense items were completed by children in grade 5 and 6. Children with scores below the range of the past tense items are presented with different types of verb items in the Verb Game, e.g., present tense items. Of the analyzed past tense items, 0.036% were made by children in kindergarten, 0.18% in grade 1, 1.3% in grade 2, 6.4% in grade 3, 15.7% in grade 4, 31.6% in grade 5, 40.2% in grade 6, and 4.6% by older children. These results are in line with the Dutch language curriculum: the past tense spelling rule is rather difficult and only explicitly taught in the later grades of primary school. Note here that we conduct item-based analyses instead of child-based analyses.

Language Sea interface

Language Sea games are presented as a sea with schools of fish, each school representing a game (e.g., an octopus for the Verb Game). By playing the games, the number of fish in the school increases (the higher the ability, the more fish in each school). For this article, we focus on the Verb Game. If a child clicks on the octopuses, the Verb game starts. Children are presented with an item and are given 20 s to answer. The item consists of a sentence in which the infinitive is provided between parentheses. The child has to inflect the infinitive correctly, such that it fits within the sentence (see Figure 2). Different inflections and tenses are asked: present tense, past participle, and participle

Figure 2. The Verb Game from Language Sea. The child has to inflect the infinitive, presented between parentheses. The cue 'vorige week' (last week) shows that a past tense is required.

adjectives. Past tense is indicated either with a cue (v.t., a common Dutch abbreviation for *verleden tijd*, past tense), or indirectly by a time cue in the sentence (such as *vorige week*, last week).

The remaining time is reflected as a row of coins in the bottom of the screen, from which a coin disappears with each passing second. Upon answering, the correct answer is shown and the child receives the number of remaining coins if the answer was correct, but loses the same number of coins if the answer was incorrect (but the total number of coins won in each game can never be below zero). There is a question mark that the child can click if (s)he does not know the answer: in this case, and also when the child does not provide an answer within the time limit, no coins are won or lost and the next item appears. The rationale of this scoring rule is explained in the next section. The coins can be used to buy virtual prizes in a trophy cabinet. The child is thus motivated to answer quickly if (s)he knows the answer, but to refrain from answering otherwise. A game ends after 15 items, but children can quit a game earlier or play the game several times.

Language Sea computer adaptive technology

By means of a computer adaptive algorithm based on Item Response Theory (IRT) modeling, in *Language Sea* every instance of a child solving an item is used to update the estimates of the ability of the player and the difficulty of the item. These estimates are then used to select the next item. The updating procedure is based on a procedure invented by Elo (1978), used to rank chess players. This method is briefly outlined here; for a full description, we refer to Klinkenberg, Straatemeier, and Van der Maas (2011), and for the mathematical foundations to Maris and Van der Maas (2012).

Ability of the child and difficulty of the item are both expressed on the same scale. The scale itself is arbitrary, but the child's expected score on a particular item can be derived from the difference between child ability and item difficulty, in a way very similar to IRT modelling (Klinkenberg et al., 2011). This expected score is somewhere between −1 (immediate wrong answer) and 1 (immediate correct answer). After the child solves the item, the expected score and the actual, obtained score are compared. The child's ability estimate is then adjusted upward if the child scored higher than expected and downward if the child scored lower than expected. In a similar fashion, the item difficulty estimate is also adjusted: downward if the child scored higher than expected and otherwise upward. New children and new items enter the system with an estimated starting value. These estimates are updated after every played trial, and they tend to approach their stable value quickly (Klinkenberg et al., 2011).

While in the original Elo (1978) method only wins, draws and losses are considered, and, similarly, standard IRT models are based on accuracy, the *Language Sea* algorithms also take response time into account. This leads to a system with strong psychometric properties (Klinkenberg et al., 2011; Maris & Van der Maas, 2012). The item selection procedure is also based on item difficulty and child ability. Based on the child's ability, the difficulty level is determined at which the child has a probability of .75 of answering correctly. The next item is sampled around this difficulty level (M = .75, SD = .10). Children can also choose to solve easier items, with an average expected probability of .90 correct, or more difficult items, with an expected probability of .60.

This method of data collection and item selection works well and, because both child ability and item difficulty are estimated on the fly, it enables analyses on the level of the items and on the level of the users (see Gierasimczuk, Van der Maas, & Raijmakers, 2013; Jansen et al., 2014; Van der Ven, Klaiber, & Van der Maas, 2016; Van der Ven, Straatemeier, Jansen, Klinkenberg, & Van der Maas, 2015; Van der Ven, Van der Maas, Straatemeier, & Jansen, 2013). In the present paper, analyses are focused on the difficulty of the items: how come children find some items more difficult than others? Because of the adaptive algorithm, the mere number of errors is not informative: more difficult items are presented to more skilled players, such that the error rate is approximately equal for all items (i.e., approximately 25%). We therefore analyzed the difficulty estimates. In addition, we compared the types of the errors that were made on different items.

Verb game items

The *Language Sea* Verb Game contains 300 regular past tense items. For the present study, verbs with stems ending in *-t* (e.g., *zetten*, to put) or *-d* (e.g., *wedden*, to bet) were not taken into account, as it is impossible to interpret the nature of these errors (see also Ernestus & Baayen, 2003). Recent loanword verbs such as *freelancen, racen, lunchen*, and *lobbyen* were also excluded. The verbs themselves as well as the past tenses proved difficult to spell correctly (*freelancete, racete, lunchte, lobbyde*). These verbs are infrequent in Dutch, even in the adult-based corpus (token frequency information is not attested for all verbs; some verbs have been introduced more recently in Dutch and are thus not always present in the existing corpora). Furthermore, the relationship between phonology, morphology, and orthography is very different for these verbs than non-loanwords. These verbs thus do not impact on regular past tense spelling in general and render different outcomes that are not relevant for the cues assessed here. The remaining 227 regular verbs (101 taking-*de* and 126 taking-*te*) were analyzed. The prefinal row of Table 2 shows an overview of the number of items, divided by type of verb; a full list of items is given in Appendix 2.

We used past tense data collected between October 2012 and March 2014. For the difficulty analyses, the difficulty of each item was extracted on a weekly basis. These difficulty estimates stabilise quite quickly; to get even more stable results, the mean difficulty of the last 30 weeks was taken. For the error analyses all errors that were made during the 79 weeks were included: 140,202 errors (in a total of 392,802 solved items; 36% errors). Note that the adaptive algorithm implies that children are only presented with (past tense) items with a difficulty that matches the child's ability.

Past tense variables

There were different variables of interest. One was labeled the IMP variable. This variable captured the number of converging cues, ranging from 2–6 (see Table 1). Table 1 shows that most cues were clearly converging or colliding, but there were some 'mixed' cues, which were converging for some verbs and colliding for others, even within a verb type. For instance, for fricatives taking-*te*, graphotactic frequency was a converging cue for item *bruiste*, with a graphotactic frequency of 89% for *st*, but a colliding cue for item *besefte*, with a graphotactic frequency of only 15% for *ft*. In these cases, this cue was assessed for convergence for each separate item (sometimes "yes" and sometimes "no").

In subsequent analyses the cues were analyzed separately, rather than being collapsed as colliding or converging. The following variables were used: verb type (consistent: nasals, liquids, glides; semi-consistent: plosives, and inconsistent: fricatives, distinguishing between fricatives taking *-te* and *-de*),

Table 2. Multiple regression analyses explaining the item difficulty of all past tense items (n = 227).

Variable	Model 1					Model 2					Model 3				
	B	*SE*	*t*	*p*	*St*[1]	*B*	*SE*	*t*	*p*	*St*[1]	*B*	*SE*	*t*	*p*	*S*[1]
Control															
Intercept	6.65	0.48	13.69	<.001	1.00	9.64	0.74	13.11	<.001	1.00	5.98	0.80	7.48	<.001	1.00
Word length	0.04	0.09	0.41	.680	1.00	0.04	0.09	0.49	.622	1.00	0.12	0.08	1.55	.123	.90
Spelling difficulty infinitive	0.31	0.09	3.61	<.001	1.00	0.33	0.08	3.97	<.001	1.00	0.32	0.07	4.53	<.001	1.00
Plural (vs singular)	0.92	0.23	3.94	<.001	.84	0.90	0.22	4.09	<.001	.84	1.09	0.19	5.76	<.001	1.00
Token frequency	−0.05	0.05	−0.99	.325	1.00	−0.08	0.05	−1.61	.113	.90	−0.06	0.04	−1.37	.172	.97
Capital	1.85	0.34	5.42	<.001	1.00	1.95	0.32	6.02	<.001	1.00	2.03	0.28	7.35	<.001	1.00
Two words	0.63	0.39	1.60	.112	.90	0.84	0.38	2.24	.023	.77	0.55	0.33	1.69	.091	.84
Past tense cue	−0.19	0.21	−0.91	.359	1.00	−0.11	0.20	−0.57	.567	1.00	−0.20	0.17	−1.23	.223	1.00
IMP															
No of converging cues						−0.57	0.11	−5.20	<.001	1.00					
Morphophonological and orthographic															
Fricatives-*te*											0.93	0.35	2.68	.008	1.00
Fricatives-*de*											1.15	0.30	3.80	<.001	1.00
Plosives-*te*											0.05	0.21	0.24	.813	1.00
Voicing consistency											2.80	2.56	1.10	.275	1.00
Voicing consistency * fricatives-*te*											−10.58	2.83	−3.74	<.001	1.00
Voicing consistency * fricatives-*de*											−4.22	2.97	−1.42	.157	.97
Voicing consistency * token frequency											0.75	0.42	1.81	.072	.81
Graphotactic frequency											−0.74	0.64	−1.10	.243	1.00
R2	.36					.42					.61				
ΔR2						.06					.19				

[1]St = stability: proportion of individual, weekly analyses in which the significance (with an alpha of .05) was the same as reported.

voicing probability (taken from Ernestus & Baayen, 2003), and graphotactic frequency, calculated on the basis of CELEX (Baayen et al., 1993).

The value of voicing probability was transformed to voicing consistency in the regression analysis. This was done as a high voicing probability should facilitate selection of-*de;* it is thus negatively related to the difficulty of -*de* items but positively to the difficulty of -*te* items. Voicing consistency is equal to voicing probability as defined by Ernestus and Baayen (2003) for verbs taking voiced allomorph-*de* (e.g., voicing probability, while voicing consistency is 1 minus voicing probability for verbs taking voiceless allomorph-*te* (e.g., vissen: voicing probability = .135; voicing consistency = 1 —.135 = .865). Higher voicing consistency is always predicted to lead to a lower difficulty and fewer allomorph errors.

Graphotactic frequency of the stem-final grapheme and the correct consonant (*t* or *d*) was determined based on all past tenses and participles from the CELEX database (Baayen et al., 1993). For instance, for the verb "*noemen*" name with past tense "*noemde*", the graphotactic frequency was (frequency of *md*/(frequency of *md* + frequency of *mt*), rendering a value of 1 (i.e., always *md*). For verbs with stems ending in "i", the frequency of digraphs *aai*(d/t), *oei*(d/t), and *ei*(d/t) was calculated.

Characteristics that are also expected to play a role are entered first as control variables. These control variables are word length in letters, spelling difficulty estimate of the infinitive, derived from the Flash Game in *Language Sea*, in which words have to be spelled that are first flashed for 600 ms, plural spelling (vs. singular), the natural logarithm of the token frequency (from CELEX; Baayen et al., 1993), as well as the task-specific cues capitalization, two word verbs (vs. one word), and time cue.

Error classification

Realisations of past tense targets in *Language Sea* consist of a wide variety of errors. All errors that had been entered at least three times were categorised as typos, morphological errors, allograph errors, vowel changes, or "other". Typo errors were minor errors we were not interested in for the purposes of the present paper: (a) an accidental insertion of a symbol (e.g., *werkte\for *werkte* worked), (b) an error in capitalization, (c) omission of a letter that was not in stem final position (*wekte), (c) doubling (*werrkte), (d) singular/plural mistake (*werkten), (e) extra word (*werkte jij—worked you), or a combination of these. If these errors were the only errors that were made, the answer was categorised as a typo.

Morphological errors consisted of answers that were either the verb stem (*werk) or incorrect inflections (such as **werken* to work, **werkt*, works). Allograph errors were those in which the incorrect allograph was selected (**werkde*, worked or **stormte*, stormed). Vowel changes consisted of answers in which a non-typo vowel change occurred while the stem consonants were correct. These vowel changes could either lead to a realization with past tense marked through an allograph (e.g., *overtijgde for *overtuigde* convinced) as well as those resembling irregularisations (e.g., **kneeg* for *knaagde* gnawed). The remaining category "other" included realizations with -d endings (e.g., **miauwd* for *miauwde*, meowed), making it impossible to decide whether it is a morphological error (attempt at producing a present tense verb with final *-d* instead of final *-t*) or a typo (omission of the final *-e*). Furthermore, it included realisations of targets with other verbs than those targeted (e.g, **smolt*, melted, for *smeulde*, smouldered), distortions (e.g., **verzwikkelde* for *verzwikte* sprained), as well as realisations with combinations of errors (e.g., **verdwenen* disappeared for *verwende*, spoiled, including a change of verb as well as a morphological error).

Analyses

The analyses consist of two parts: difficulty analyses and error analyses.

Difficulty analyses

In the first part, regression analyses show which characteristics predict the difficulty estimates of the items. We analyzed the variables described in the Past Tense Variables section: in Model 1 only the control variables, in Model 2 we added the IMP variable, and in Model 3 we included the morphophonological and orthographical variables of interest: consonant type, voicing consistency, and graphotactic frequency.

Error analyses

The second part assesses the error patterns in relation to final consonants, divided into consistent, (semi-consistent) plosives, and (inconsistent) fricatives. Chi square analyses were performed to assess differences in error patterns between these consonant types. As a measure of effect size, odds ratios were used. These can be used to compare whether a certain type of error is made equally often for two consonant types. The odds of a certain error ($\frac{number\ of\ errors\ in\ target\ category}{number\ of\ other\ errors}$) were determined for these two consonant types. Subsequently, the ratio of these odds was determined. An odds ratio of 1 (95% confidence interval including 1) implies that for both consonant groups the odds of a certain error type was the same, and thus no effect of this error type. If the odds ratio deviates from 1, this error type was more prevalent for one consonant type than for the other; the larger the deviation from 1, the larger the effect.[1] Since we distinguished three consonant types, all pairs of two were compared separately (consistent vs. plosives, plosives vs. fricatives, and consistent vs. fricatives).

Results

Difficulty analyses

In order to assess the strength of factors contributing to relative item difficulty of past tense productions, a hierarchical regression analysis was conducted. The results are presented in

[1]Note that the odds ratio does *not* give the difference in probability of the target error between two groups, since the odds do not equal the probability of this error (that would be $\frac{number\ of\ target\ errors}{total\ number\ of\ errors}$).

Table 2. In Model 1, the control variables were entered. Model 1 explained a significant proportion of variance, $R2 = .36$, $F(7, 219) = 17.34$, $p < .001$. The stability column in Table 2 shows in how many individual weekly analyses the results were the same; as stability is high for every predictor, the results are highly reliable. Difficulty of the spelling of the infinitive, a plural verb answer instead of singular, and having to capitalise the verb turned out to be significant control variable predictors (also in Models 2 and 3).

In Model 2, the number of converging cues was added to test the IMP hypothesis. This was a significant predictor: the more converging cues, the lower the difficulty of the item. The addition led to an increase in explained variance with 6%, $R^2 = .42$, and a significant improvement in model fit, $F(1,218) = 27.03$, $p < .001$.

In Model 3, the importance of the separate cues was assessed: the "number of converging cues" variable was replaced by variables representing the different cues, which were further refined according to orthography and morphophonology. In addition, interaction terms were added to investigate whether the effect of voicing consistency differed between consonant types and varied as a function of token frequency. As an illustration, Figure 3 shows the observed and predicted difficulty estimates of five representative items from each of the four consonant categories. The horizontal axis shows the observed difficulty estimates; the vertical axis shows the difficulty estimates as predicted by Model 2. This model, displayed in Table 2, explained 19% more variance than Model 2, a significant improvement, $F(4, 211) = 10.45$, $p < .001$. The first three morphophonological variables are the three consonant types: these indicate the difference in difficulty between the respective consonant category and the reference category: the consistent consonant types (glides, nasals, and liquids). As expected, both types of fricatives (inconsistent) were more difficult than consistent consonants, whereas plosives (semi-consistent) did not differ significantly from the consistent verbs. The main effect of voicing consistency (which reflects the effect of voicing consistency for the plosives) was not significant. However, the interaction term of voicing consistency x fricatives taking-*te* was significant, meaning that verbs with stems ending on a fricative taking *-te* were easier if voicing consistency was high. As in Model 1, stability was high. There was no effect of graphotactic frequency (note that further interaction terms with graphotactic frequency could not be included because of multicollinearity).

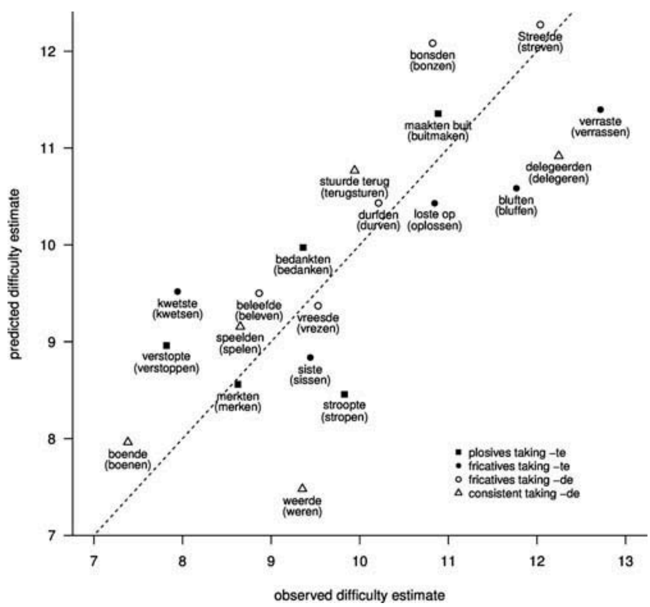

Figure 3. The observed and predicted difficulty estimate of five representative items from each of the four consonant categories (predictions are based on Model 3, displayed in Table 2). The correct answer and the infinitive (between parentheses) are displayed.

These findings thus show that when producing past tenses, children displayed sensitivity to morphophonological characteristics and frequency distributions of the verb they were inflecting: verbs with ambiguities as a consequence of devoicing were more difficult than consistent verbs, especially when the correct inflection of the verb did not match the voicing probability of the rhyme of the stem. The findings show that different cues are used in item past tense inflection. Next, we will zoom in on the types of errors made in the task and investigate whether the error patterns can also be related to orthomorphophonological characteristics.

Past tense errors

In the first error analysis, the errors patterns made on the different types of verbs: consistent verbs (all taking *-de*)), semi-consistent and inconsistent verbs, were compared. The relative frequencies of these errors are displayed in Figure 4.

The overall error pattern differed between the three different types of verbs, $\chi^2(8) = 15{,}550.56$, $p < .001$. A series of chi square analyses was performed to flesh out the association by error type (see Table 3). The column "order of frequency" shows for which verb type the error was made most often (see also Figure 4). The frequencies of all error types differed significantly between the different verb categories. Odds ratios show the size of the difference in occurrence between each combination of two verb categories. These differences were always significant, except that typing errors were made equally often on semi-consistent and consistent verbs. Very large effects are present in the wrong allograph errors: the ratio of wrong allographs to all other types of errors is 39.04 times as large for inconsistent as for consistent verbs and 14.85 times as large for inconsistent as for semi-inconsistent

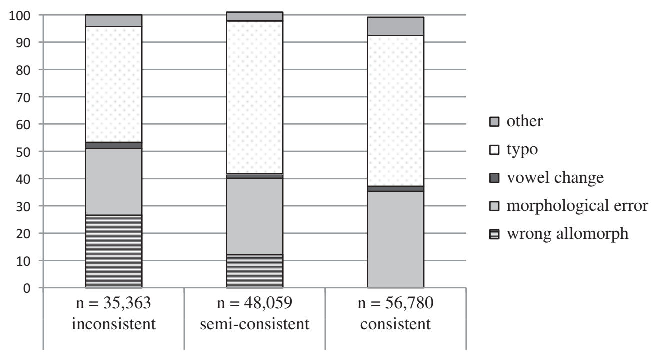

Figure 4. Plots showing the frequencies of the different error types by verb consonant category.

Table 3. Odds ratios of the differences in error frequencies between past tense verb types.

			Odds ratio (95% C.I.)		
Error type	χ2	Order of frequency	inconsistent vs. semi-consistent	semi-consistent consistent	Inconsistent vs. consistent
Wrong allomorph	14,483.53**	inconsistent > semi-consistent > consistent	2.63** (2.53–2.73)	14.85** (13.56–16.29)	39.04** (35.69–42.74)
Morphological	1,385.05**	consistent > semi-consistent > inconsistent	1.20** (1.17–1.24)	1.40** (1.37–1.44)	1.69** (1.64–1.74)
Vowel change	55.70**	inconsistent > consistent > semi-consistent	1.45** (1.31–1.60)	1.17** (1.06–1.28)	1.24** (1.13–1.36)
Typing error	1727.25**	consistent = semi-consistent > inconsistent	1.68** (1.63–1.71)	1.01 (0.98–1.03)	1.68** (1.63–1.72)
Other	723.92**	consistent > inconsistent > semi-consistent	1.36** (1.26–1.46)	2.18** (2.05–2.32)	1.61** (1.51–1.71)

** $p < .001$

verbs. In other words, wrong allograph errors are much more likely to occur for inconsistent than for (semi-) consistent verbs.

In the second error analysis, the errors made on inconsistent verbs (fricatives) taking-*de* and-*te* and the semi-inconsistent verbs (plosives) were broken down by voicing probability (the original values as defined by Ernestus & Baayen, 2003, so not voicing consistency). The results are shown in Figure 5. The plots show that for the fricatives taking-*te*, the wrong allograph occurs far more frequently than for both other verb types, and there is a steep increase in the frequency of selection of a wrong (i.e., voiced) allograph as voicing probability increases. For fricatives taking-*de*, the reverse pattern is shown: the selection of a wrong (i.e., unvoiced) allograph decreases as voicing probability increases. For plosives no pattern can be discerned, but voicing probability is always low. The findings show that children are sensitive to voicing probability, and are more likely to choose an incorrect allograph if voicing probability is inconsistent with the actual voicing value. However, the graph also shows a high degree of wrong allograph selection for the fricatives taking-*te* with a voicing probability of .081. These verbs ending in an *f* take a low graphotactic frequency of *ft*, (see also Appendix 2) pushing towards *fd*. The relationship between voicing probability, graphotactic frequency, and errors was tested further. For each verb, the logarithm of the odds of a wrong allograph error was determined. In a regression analysis, presented in Table 4, Model 1, voicing consistency and graphotactic frequency were entered as predictors of this logarithm of odds. Together, they predicted 40% of the variance in degree of wrong allograph selection between verbs, $F(2,224) = 74.92$, $p < .001$, but only voicing consistency was a significant predictor. Thus, for verbs with a voicing probability that was not consistent with the actual correct allograph, children were more likely to form the past tense using the wrong allograph.

Model 2 showed that this pattern differed per type of verb, $F(7,219) = 106.1$, $p < .001$. Interaction effects with fricatives taking -*de* were initially also included, but had to be removed because of high multicollinearity (these were not significant). The interaction effects show that for fricatives taking-*te*, the effect of voicing consistency on allograph selection is strong, and an effect of graphotactic frequency was also present for these verbs. For the plosives and the fricatives taking-*de*, voicing consistency did not contribute to the degree of wrong allograph selection, although fricatives taking-*de* and plosives had a higher degree of wrong allograph selection across the board compared to the reference category (the consistent verbs).

Past tense errors in inconsistent verbs (fricatives)

To further ascertain whether the past tense allograph errors were driven by morphophonological and graphotactic sensitivity, additional error analyses were conducted for fricatives taking-*de*. As these verbs have underlying stem endings on -*v* and -*z* but spelling of -*f* or -*s* in the past tense (*erfde* inherited, *kneusde* bruised), the relationship between spelled voicing value of the stem-final consonant and the allograph is relevant.

Three error types were compared: (a) both incorrectly spelled stem-final consonant and incorrect allograph (e.g., **ervte* for *erfde* or **kneuzte* for *kneusde*), in which graphotactic frequency of incorrect *vt* and *zt* is 0; (b) incorrectly spelled stem-final consonant but correct allograph (**ervde* or **kneuzde*), with graphotactic frequencies for *vd* and *zd* of 0; and (c) correctly spelled stem-final consonant but incorrect allograph (**erfte* or **kneuste*)[2], with a graphotactic relative frequency *ft* of 0.11 and *st* of 0.89. In the previous analyses, type (a) and (c) were categorised as wrong allograph errors, and type (b) as a typing error. Error type (a) reflects a morphophonological as well as orthographical error. Error type (b) refers to incorrect orthography, but correct morphophonology, as the pronunciation of the target would indeed be /vd/and /zd/. Similar to error type (b), errors of type (c) also show alignment between the stem-final consonant and the allograph; the spelling of the stem-final consonant is correct, but allograph selection is not.

[2]In the previous analyses, type (a) and (c) were categorized as wrong allograph errors, and type (b) as a typing error.

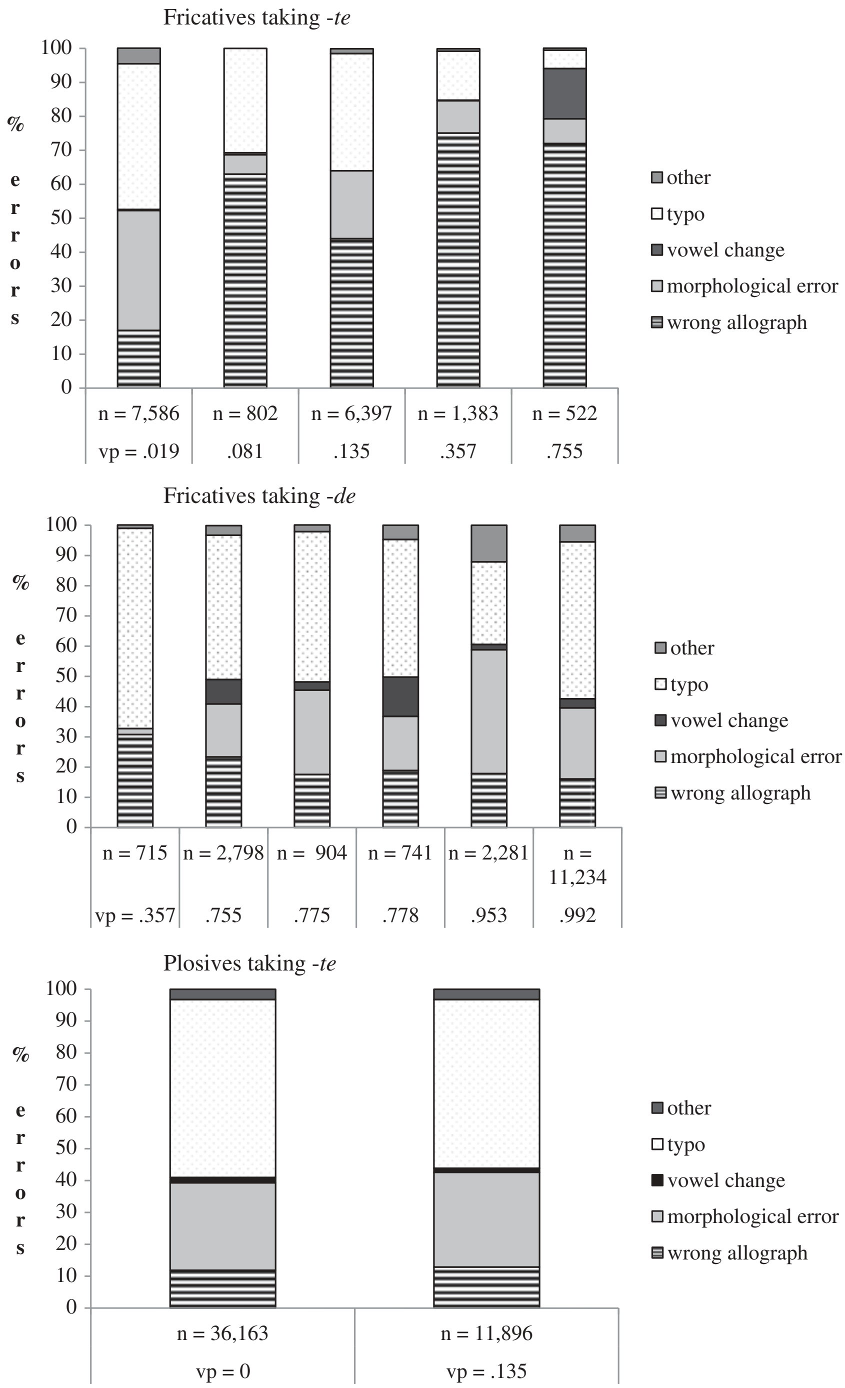

Figure 5. The distribution of the different errors for fricatives taking *-te* (upper panel) and *-de* (middle panel) and plosives (lower panel), broken down by voicing probability.

Table 4. Regression analyses predicting the degree of wrong allograph selection.

	Model 1				Model 2			
Variable	*B*	*SE*	*t*	*p*	*B*	*SE*	*T*	*p*
Intercept	0.15	0.08	1.81	.072	0.04	0.09	0.44	.660
Voicing consistency	−1.65	0.16	−10.50	<.001	−0.23	0.18	−1.26	.208
Graphotactic frequency	0.02	0.09	0.23	.816	−0.02	0.09	−0.22	.827
Fricatives-*te*					1.04	0.14	7.69	<.001
Fricatives-*de*					0.17	0.03	4.80	<.001
Plosives-*te*					0.11	0.02	4.43	<.001
Voicing consistency * fricatives-*te*					−1.74	0.23	−7.44	<.001
Graphotactic frequency * fricatives-*te*					−0.70	0.15	−4.55	<.001
R^2	.40				.77			
ΔR^2					.37			

Error type (a) occurred 116 times (2%), type (b) was made 2962 times (58%), and type (c) was made 2064 times (40%). Binomial tests showed that error types (b) and (c) were made significantly more often than type (a), $p < .001$, confirming that in their errors children displayed morphophonological and orthographic sensitivity: they preferred congruence in voicing of stem and allograph. Type (b) occurred significantly more often than (c), $p < .001$: children preferred a match consistent with phonology that violates Dutch spelling rules over a match with correct stem-final consonant spelling. In these instances they ignore the information provided by graphotactic frequency (that *vd* and *zd* do not occur).

In contrast, the voicing value of fricative *-te* verbs is the same in infinitive and stem (e.g. *sissen–sis–siste*, to hiss-hiss hissed, *beseffen-besef-besefte*) and is consistent in spelling. The graphotactic frequency of *st* is high (0.89), that of *ft* is low (0.11). These cues render it unlikely that errors such as **sizde* and **besevde*, i.e., incorrectly spelled stem-final consonant and incorrect allograph will occur, or **sizte* and **besevte*, i.e., incorrectly spelled stem-final consonant but correct allograph. The only error type that could be anticipated is**sisde* and **besefde*, i.e., correctly spelled stem-final consonant but incorrect allograph. Indeed, only the third error type was present in the data (3396 times).

Discussion

The Integration of Multiple Patterns hypothesis (IMP, Treiman & Kessler, 2014) puts forward that learning to spell is dependent on different patterns, providing cues for spelling. Words are easier to spell when these cues converge on the correct answer than when they collide. In this study, the IMP was evaluated in light of the complex spelling pattern of Dutch past tense, which is notoriously difficult and problematic in the Dutch education system and beyond (e.g., Assink, 1987; Bonset & Hoogeveen, 2009; DeSchryver et al., 2013; Ernestus & Baayen, 2001; Neijt & Schreuder, 2007; Zuidema, 1988). We assessed to what extent morphophonological and orthographic cues contribute to Dutch past tense item difficulty through item difficulty and error analyses. Written past tenses of Dutch primary school-aged children playing the *Verb Game*, part of a Dutch computer-adaptive learning system, were analyzed.

After controlling for a number of relevant control variables, we found that for the specific morphophonological and orthographic cues, the number of converging cues was a significant predictor of item difficulty. This was predicted by the IMP. More converging cues led to a lower item difficulty. However, we also found that not every cue was equally important, and that the importance of the cues differed based on morphophonology. As expected, stem consonant was found to contribute to item difficulty: fricatives were more difficult than other consonants. This difficulty was further affected by and interacted with voicing probability (measured through voicing consistency): especially fricative-*te* targets were difficult when the voicing consistency value leans towards a-*de*. These findings thus show that children are sensitive to (implicit) morphophonological patterns in their production of past tenses. Contrary to expectations based on the IMP, graphotactic frequency did not contribute significantly to item difficulty. This might be due to the fact that voicing consistency and graphotactic frequency often point to the same correct (e.g., in verbs such as *rennen*-run, *rende*) or incorrect allograph (e.g., in verbs such as *bonzen*-pound both point towards incorrect *st*, **bonste*). It should be noted here that our measure of graphotactic frequency is narrow, as we only looked at bigram frequencies of the final stem-grapheme and the onset-allograph grapheme. It could be the case that children also apply frequencies from larger chunks, such as the syllable rhyme. However, if these had been tallied, the overlap between voicing probability and the orthographic measure would have been substantial.

The findings on item difficulty warranted error analyses divided in targets differing in voicing consistency. Regarding the error patterns, typing errors (with correct allograph selection), morphological errors, and allograph errors occurred most often overall, whereas vowel change errors and "other" errors occurred relatively infrequently. There was a high occurrence of typing errors. This finding matches that of the item difficulty regression analyses, in which the spelling convention of capitalisation influenced item difficulty. Overall, the error pattern differed between the consistent (nasals, glides, liquids), semi-consistent (plosives), and inconsistent (fricatives) verbs. The consistent verbs were characterized by typing and morphological errors but hardly any allograph errors. In contrast, for inconsistent verbs, the allograph errors were highly frequent, and for the semi-consistent verbs the frequency was between these extremes. Regression analyses showed that voicing consistency and graphotactic frequency predicted 40% of the variance in wrong allograph selection. The higher difficulty of inconsistent verbs attested in the regression analyses was accompanied by a higher degree of wrong allograph selection for these more difficult verbs. These findings relate to adult data (DeSchryver et al., 2013; Ernestus & Baayen, 2001, 2003, 2004) and oral past tense inflection data of Dutch children (Kerkhoff, de Bree, Hoeben, & Vreugdenhil, 2014; Rispens & de Bree, 2014), which show that allograph errors predominantly occur for fricative verbs.

Within the fricatives, error patterns differed between those taking the-*de* and-*te* allograph. In both, typing errors were equally frequent, but the occurrence of allograph errors differed. The allograph errors were much more dominant in the fricative verbs taking-*te*. Regression analyses showed that allograph errors in fricative-*te* verbs were influenced by voicing probability and graphotactic frequency. Increased voicing probability (i.e. tending towards-*de*), led to more fricative-*te* verbs being produced with-*de*. This finding indicates sensitivity to phonological cues of the rhyme structure and the stem value in past tense morphology. This is endorsed by the finding that allograph errors often co-occurred with voicing changes in the stem-final grapheme (e.g., **ervde* instead of *erfde*-inherited). In other words, graphotactic frequency information (absence of *vd, zd, vt*, and *zt* in Dutch past tense and participles) was sometimes ignored in favor of morphophonology. However, allograph errors for fricative-*te* verbs were also related to graphotactic frequency, as the *besefde* for *besefte*-realised errors show; graphotactic frequency of *fd* (85%) is dominant in these errors.

In conclusion, the results show that the past tense spelling rule children are taught is not applied consistently, as there are many errors. They also indicate that there is no hypercorrection to *d*: our data show that-*de* errors can occur, but that the occurrence of allograph errors is dependent on many different cues and does not occur across-the-board. Furthermore, hypercorrection is not likely to cause alignment of the final stem consonant and the allomorph (e.g., **eizde*), as the allomorph-*de* would then be applied regardless of previous consonants.

Although the variables included were able to account for item difficulty to a sizeable extent, not all variance was captured. There are thus other variables that also contribute to item difficulty. Similarly, although allograph errors can generally be interpreted on the basis of morphophonological and orthographical patterns, there are some exceptions. Stem-final *k* verbs (e.g., *boeken*-book), for instance, have a voicing probability of 0 and a graphotactic frequency of 1, i.e., both pushing towards-*te*. Nevertheless, allograph errors do occur, although to a much lesser extent than when cues clash with the correct answer. Additional variables thus need to be introduced. The data thus provide support for a spelling model in which integration of orthographical, phonological, and morphological knowledge is proposed, specifically the IMP: The explicit rule of past tense spelling is often overruled in favor of implicit statistical information, such as voicing probability and bigram grapheme frequency. These findings thus call for more in-depth (linguistic) analysis of spelling acquisition (see, e.g., Ravid, 2012; Turnbull, Deacon, & Kay-Raining Bird, 2011). They also show that the *number* of converging and colliding cues do not account for all findings, but that the *values* and dominance of these cues play a decisive role. This is an important finding and can be used to specify the IMP.

In terms of spelling didactics, the findings, which show that the Dutch past tense spelling rule is difficult, reiterate the need for spelling strategies that aid past tense spelling, as has been claimed by others (Assink, 1987; DeSchryver et al., 2013). The spelling rule guides children to the voicing value of the infinitive and thus to the correct allograph (voiceless stem takes-*te*, voiced takes-*de*). However, instead of the explicit rule, children rely on implicit cues such as voicing probability and graphotactic frequency. These do point toward the correct allograph in consistent verbs, but do not in semi-consistent and inconsistent verbs. Instruction should thus focus on ensuring that children know the explicit spelling rule. More importantly, children should learn to recognise in which cases they cannot rely on their implicit knowledge and have to consciously apply the rule: for the verbs with stems ending in voiced [b], [v], and [z], and voiceless [f] and [s]. A new mnemonic with only these consonants might be helpful. This can be accompanied by exercises with repeated dictations in which low frequent and pseudoverbs need to be inflected; on the basis of the errors that will almost certainly occur, the rule can be presented. Teaching the rule explicitly by integrating the implicit knowledge that children use and should learn to ignore with the explicit rule might lead to improvement in past tense spelling.

This approach seems might stand a better chance than a proposal to change the spelling rules of Dutch, to allow different spelling options of the inconsistent verbs (DeSchryver et al., 2013). It is also more feasible than one in which rote learning of all the past tense of the verbs needs to take place. Rote learning is not possible, as there are simply too many verbs, many of which children do not know yet at the start of past tense spelling learning. Neither is it necessary, since children do not often err in the past tense of consistent verbs. The opposite approach, an analogy-based method, in which children are taught to match the verb that needs to be inflected to one of six prototypical verbs, has not been found to be successful (e.g. Assink, 1987; Zuidema, 1988). Past tense spelling did not improve strongly, possibly due to the fact children do not understand the principle behind the prototypes. In other words, knowledge remained too implicit. However, in the method we propose, analogies would not be used as an alternative to an algorithmic rule but rather as support to illustrate in which cases intuitions conflict with the rule. They may serve as a means to add to children's understanding. The method we propose combines explicit and implicit (oral and written) knowledge and entails substantial feedback.

The findings of our study show that multiple probabilistic cues influence spelling. These oral (morphophonology) and written (graphotactic frequency) cues contribute to spelling outcomes and can even overrule spelling instruction. As a result, the findings have both practical and theoretical implications. On the practical level, the results can be used to enhance teachers' awareness of the mechanisms behind (in)correct spelling, such that they can focus their instruction especially on the difficult cases. On a theoretical level, the results show a connection to oral language acquisition (e.g. Turnbull et al., 2011), as allograph errors mainly occur in inflecting fricative verbs for past tense (e.g., Kerkhoff et al., 2014; Rispens & de Bree, 2014). Furthermore, the findings relate to assumptions that component skills are used and integrated from an early age onwards, and coincide with theories of language acquisition that propose that language learning is probabilistic in nature and dependent on distributional statistical learning (e.g., Bybee, 2007; Pierrehumbert, 2003).

The findings on Dutch past tense inflection proved a good test case for the IMP and can be used to refine it the IMP. The IMP as such is not very detailed in its assumptions. Our results point towards the merits of the IMP, but also show that not all cues are equally important. Future studies could look into different spelling patterns and in different orthographies to further evaluate the IMP as well as gain insight in the relative contribution of the different morphophonological and orthographic cues. This is relevant for all spelling patterns that demand the interaction between morphology, phonology, and orthography, such as English past tense, in which spelling of a morpheme deviates from the form expected on the basis of other phonological or orthographic patterns. Next to assessing patterns within different languages/orthographies, cross-linguistic comparisons could also be made. Spelling of double consonants, for instance, has been found to be difficult for Dutch (e.g., Noteboom & Reitsma, 2007), but also Norwegian (Uppstad & Solheim, 2007) and Danish (Juul & Sigurdsson, 2005). A cross-linguistic comparison of the influence of the different cues contributing to (un)successful spelling of the double consonants would specify the IMP and establish the impact of language typology on the interaction between morphology, phonology, and orthography (see, for instance, Gillis & Ravid, 2006; for a comparison between Dutch and Hebrew with respect to homophonous graphemes, and Landerl & Reitsma, 2005; on a comparison between German and Dutch vowel duration spelling) and ultimately contribute to spelling awareness and teaching. Thus, more insight into specific spelling patterns, as well as cross-linguistic comparison of spelling acquisition might specify the role that different cues play for spelling development, which would benefit spelling awareness and instruction.

There are two methodological considerations that need to be taken into account. The first is that an adult-based corpus was used for calculating verb token frequencies and graphotactic frequencies. Storkel and Hoover (2010) compared phonotactic probability and neighbourhood density values in child and adult corpora. They found significant correlations between the values, indicative of similar gross patterns. However, they also found differences between child and adult values and distributions. This might also be the case for token and graphotactic frequency in this study. Availability of child-based corpora is thus welcome.

The second consideration is that data of the present study came from a computer-adaptive programme. As data collection takes place online, it allows analysis of a very large and comprehensive sample, containing a much larger age and ability range than designs generally used in spelling research. The online setting, however, reduces the degree of control over playing frequency and circumstances. At the same time, this setting is more ecologically valid than a lab setting; children's spelling is measured while they are practicing during their daily routine.

The computer-adaptive item selection procedure implies that each child solves different items. For the present study this was not a problem, but other scientific questions cannot be answered with this approach and still require a design of targeted small-scale experiments. The approach used in the present study should therefore be considered a valuable addition to the existing body of research methods. In summary, in line with the IMP, this study showed that written past tense production is driven by different linguistic and orthographic cues. They also indicate that instruction of past tense rules has not been successful across the board yet. Children seem to have acquired the general rule of past tense production and are sensitive to phonological cues. However, in order to inflect past tenses correctly, they need to know the voicing values and overrule voicing probability and graphotactic frequency values in inconsistent verbs. On the basis of these findings and the IMP framework, we plea for a teaching approach in which the explicit past tense spelling rule is connected to the implicit knowledge children use.

References

Assink, E. M. H. (1987). Algorithms in spelling instruction: The orthography of Dutch verbs. *Journal of Educational Psychology*, *79*, 228–235. doi:10.1037/0022-0663.79.3.228

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database. CD-ROM*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bonset, H., & Hoogeveen, M. (2009). *Spelling in het basisonderwijs. Een inventarisatie van empirisch onderzoek [Spelling in primary education. An inventory of empirical research]*. Enschede, SLO. Retrieved from http://www.slo.nl/organisatie/recentepublicaties/00103/

Booij, G. (1995). *The phonology of Dutch*. Oxford, UK: Clarendon Press.
Bosman, A. M. T., Vonk, W., & Van Zwam, M. (2006). Spelling consistency affects reading in young Dutch readers with and without dyslexia. *Annals of Dyslexia*, *56*, 271–299. doi:10.1007/s11881-006-0012-4
Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford, UK: Oxford University Press.
Deacon, H. S., Leblanc, D., & Sabourin, C. (2011). When cues collide: Children's sensitivity to letter- and meaning-patterns in spelling words in English. *Journal of Child Language*, *38*, 809–827. doi:10.1017/S0305000910000322
DeSchryver, J., Neijt, A., Ghesquière, P., & Ernestus, M. (2013). Zij surfde, maar hij durfte niet. De spellingproblematiek van de zwakke verleden tijd in Nederland en Vlaanderen [She surfed, but he dared not. The spelling difficulties of the weak past tense in The Netherlands and Flanders]. *Dutch Journal of Applied Linguistics*, *2*, 133–151. doi:10.1075/dujal.2.2.01de
Elo, A. (1978). *The rating of chess players, past and present*. New York, NY: Arco Publishers.
Ernestus, M., & Baayen, H. (2001). Choosing between the Dutch past tense suffixes -te and -de. In T. Van der Woude, & H. Broekhuis (Eds.), *Linguistics in the Netherlands* (pp. 77–88). Mahwah, NJ: Lawrence Erlbaum Associates.
Ernestus, M., & Baayen, H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, *79*, 5–38. doi:10.1353/lan.2003.0076
Ernestus, M., & Baayen, H. (2004). Analogical effects in regular past tense production in Dutch. *Linguistics*, *42*, 873–903. doi:10.1515/ling.2004.031
Gierasimczuk, N., Van der Maas, H. L., & Raijmakers, M. E. (2013). An analytic tableaux model for deductive mastermind empirically tested with a massively used online learning system. *Journal of Logic, Language and Information*, *22*, 297–314. doi:10.1007/s10849-013-9177-5
Gillis, S., & Ravid, D. (2006). Typological effects on spelling development: A crosslinguistic study of Hebrew and Dutch. *Journal of Child Language*, *33*, 621–659. doi:10.1017/S0305000906007434
Jansen, B. R. J., Hofman, A. D., Straatemeier, M., Bers, B. M. C. W., Raijmakers, M. E., & Van der Maas, H. L. J. (2014). The role of pattern recognition in children's exact enumeration of small numbers. *British Journal of Developmental Psychology*, *32*, 178–194. doi:10.1111/bjdp.12032
Juul, H., & Sigurdsson, B. (2005). Orthography as a handicap? A direct comparison of spelling acquisition in Danish and Icelandic. *Scandinavian Journal of Psychology*, *46*, 263–272. doi:10.1111/sjop.2005.46.issue-3
Kemp, N., & Bryant, P. (2003). Do bees buzz? Rule-based and frequency-based knowledge in learning to spell plural –s. *Child Development*, *74*, 63–74. doi:10.1111/1467-8624.00521
Kerkhoff, A., De Bree, E., & Buesseler, H. A. (2014). Saying 't' and writing 'd'. In R. Kager, J. Grijzenhout, & K. Sebregts (Eds), *Where the principles fail: A festschrift for Wim Zonneveld on the occasion of his 64th birthday* (pp. 205–214). Utrecht, The Netherlands: UiL OTS.
Kerkhoff, A., De Bree, E., Hoeben, I., & Vreugdenhil, A. (2014). De invloed van orthografie op verleden tijdsvorming door zwakke lezers [The influence of orthography on past tense formation by poor readers]. *Stem-, Spraak-, En Taalpathologie*, *19*, 1–18. doi:32.8310/01/1914-1
Klinkenberg, S., Straatemeier, M., & Van der Maas, H. L. J. (2011). Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, *57*, 1813–1824. doi:10.1016/j.compedu.2011.02.003
Landerl, K., & Reitsma, P. (2005). Phonological and morphological consistency in the acquisition of vowel duration spelling in Dutch and German. *Journal of Experimental Child Psychology*, *92*, 22–44. doi:10.1016/j.jecp.2005.04.005
Maris, G., & Van der Maas, H. (2012). Speed–accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika*, *77*, 615–633. doi:10.1007/s11336-012-9288-y
Matthews, D. E., & Theakston, A. (2006). Errors of omission in English-speaking children's production of plurals and the past tense: The effects of frequency, phonology, and competition. *Cognitive Science*, *30*, 1027–1052. doi:10.1207/s15516709cog0000_66
Neijt, A., & Schreuder, R. (2007). Asymmetrical phoneme-grapheme mapping of coronal plosives in Dutch. *Written Language & Literacy*, *10*, 219–234.
Noteboom, A., & Reitsma, P. (2007). Spelling Dutch doublets: Children's learning of a phonological and morphological spelling rule. *Scientific Studies of Reading*, *11*, 2, 133–150. doi:10.1080/10888430709336556
Nunes, T., Bryant, P., & Bindman, M. (1997). Morphological spelling strategies: Developmental stages and processes. *Developmental Psychology*, *33*, 637–649. doi:10.1037/0012-1649.33.4.637
Pacton, S., & Deacon, S. H. (2008). The timing and mechanisms of children's use of morphological information in spelling: A review of evidence from French and English. *Cognitive Development*, *23*, 339–359. doi:10.1016/j.cogdev.2007.09.004
Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Development*, *76*, 324–339. doi:10.1111/j.1467-8624.2005.00848_a.x
Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*, 115–154. doi:10.1177/00238309030460020501
Pollo, T. C., Treiman, R., & Kessler, B. (2008). Three perspectives on spelling development. In E. L. Grigorenko, & A. J. Naples (Eds.), *Single-word reading: Behavioral and biological perspectives* (pp. 175–189). New York, NY: Erlbaum.
Ravid, D. D. (2012). *Spelling morphology: The psycholinguistics of Hebrew spelling*. New York, NY: Springer.

Rispens, J. E., & de Bree, E. (2014). Past tense productivity in Dutch children with SLI: The role of phonology and frequency. *Journal of Child Language*, *41*, 200–225. doi:10.1017/S0305000912000542

Storkel, H. L., & Hoover, J. R. (2010). An on-line calculator to compute phonotactic probability and neighborhood density based on child corpora of spoken American English. *Behavioral Research Methods*, *42*, 497–506. doi:10.3758/BRM.42.2.497

Taalzee [Language Sea] [Computer software]. Amsterdam, The Netherlands: Oefenweb.

Templeton, S., & Morris, D. (2000). Spelling. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 525–543). Mahwah, NJ: Lawrence Erlbaum Associates.

Treiman, R., & Cassar, M. (1996). Effects of morphology on children's spelling of final consonant clusters. *Journal of Experimental Child Psychology*, *63*, 141–170. doi:10.1006/jecp.1996.0045

Treiman, R., & Cassar, M. (1997). Spelling acquisition in English. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 61–80). Mahwah, NJ: Lawrence Erlbaum Associates.

Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology*, *98*, 642–652. doi:10.1037/0022-0663.98.3.642

Treiman, R., & Kessler, B. (2014). *How children learn to write words*. Oxford, UK: Oxford University Press.

Trommelen, M., & Zonneveld, W. (1979). *Inleiding in de Generatieve Fonologie [Introduction to generative phonology]*. Muiderberg, The Netherlands: Coutinho.

Turnbull, K., Deacon, S. H., & Kay-Raining Bird, E. (2011). Mastering inflectional suffixes: A longitudinal study of beginning writers' spellings. *Journal of Child Language*, *38*, 533–553. doi:10.1017/S030500091000022X

Uppstad, P. H., & Solheim, O. J. (2007). Aspects of fluency in writing. *Journal of Psycho-linguistic Research*, *36*, 79–89. doi:10.1007/s10936-006-9034-7

Van der Ven, S. H. G., Klaiber, J. D., & Van der Maas, H. L. J. (2016). Four and twenty blackbirds: How transcoding ability mediates the relationship between visuospatial working memory and math in a language with inversion. *Educational Psychology*, 1–24. doi:10.1080/01443410.2016.1150421

Van der Ven, S. H. G., Straatemeier, M., Jansen, B. R. J., Klinkenberg, S. R., & Van der Maas, H. L. J. (2015). Learning multiplication: An integrated analysis of the multiplication ability of primary school children and the difficulty of single digit and multidigit multiplication problems. *Learning and Individual Differences*, *43*, 48–62. doi:10.1016/j.lindif.2015.08.013

Van der Ven, S. H. G., Van der Maas, H. L. J., Straatemeier, M., & Jansen, B. R. J. (2013). Visuospatial working memory and mathematical ability at different ages throughout primary school. *Learning and Individual Differences*, *27*, 182–192. doi:10.1016/j.lindif.2013.09.003

Zuidema, J. J. (1988). *Efficiënt spellingonderwijs: Een leer- en expertmodel voor het spellen [Efficient spelling education: A learning- and expert model of spelling]*. Rijksuniversiteit Utrecht/Amersfoort, Netherlands: Acco.

Appendix 1

Characteristics of Dutch verbs taking-de allograph in past tense inflection

Allograph	-de(n)										
Verb category	Semi-consistent		Inconsistent			Consistent					
Articulation	Voiced Plosives		Voiced Fricatives			(Voiced) Glides		(Voiced) Liquids		(Voiced) Nasals	
Infinitive											
grapheme	b	d	g	v	z	[aa]i	w	l	r	m	n
phoneme	[b]	[d]	[ɣ]	[v]	[z]	[j]	[w]	[l]	[r]	[m]	[n]
Examples											
Infinitive	tobben	wedden	leggen	leven	reizen	aaien	duwen	bellen	sparen	noemen	rennen
Stem	tob	wed	leg	leef	reis	aai	duw	bel	spaarde	noem	ren
Past tense (spoken)	/tɔbdə/	/wƐdə/	/lƐɣdə/	/levdə/	/rƐizdə/	/aijdə/	/dywdə/	/bƐldə/	/spardə/	/numdə/	/rƐndə/
Past tense (written)	tobde	wedde	legde	leefde	reisde	aaide	duwde	belde	spaarde	noemde	rende
Affected by devoicing ?											
Spoken	Yes [p]	Yes [t]	Yes[1] [χ]	Yes[1] [f]	Yes[1] [s]	No	No	No	No	No	No
Written	No	No	No	Yes[2] (f)	Yes[2] (s)	No	No	No	No	No	No

(*Continued*)

(Continued).

Allograph	-de(n)										
Verb category	Semi-consistent		Inconsistent			Consistent					
Articulation	Voiced Plosives		Voiced Fricatives			(Voiced) Glides		(Voiced) Liquids		(Voiced) Nasals	
Ambiguity in past tense formation based on stem?											
Spoken	Yes	Yes	Yes [χ]	Yes [f]	Yes [s]	No	No	No	No	No	No
Written	No	No	No	Yes (f)	Yes (s)	No	No	No	No	No	No
Number of items in *Language Sea*	0	na[3]	9	22	10	8	4	25	34	5	9
Number of incorrect realisations available for error analysis	0	na[3]	5,062	10,098	3,513	7,971	3,624	16,804	16,175	4,051	8,155
Category in study	Inconsistent		Semi-consistent			Consistent					

[1]infinitive and past tense: voiced. Stem: voiceless.
[2]infinitive: voiced. Stem and past tense: voiceless.
[3]/d/items are excluded in the analyses as these data are not informative for our hypotheses.

Characteristics of Dutch verbs taking-te allograph for past tense inflection

	-te(n)						
	Semi-consistent			Consistent			
Allograph Verb type Articulation	Voiceless Plosives			Voiceless Fricatives			
Infinitive							
grapheme	p	t	k	ch	f	s	k
phoneme	[p]	[t]	[k]	[χ]	[f]	[s]	[k]
Examples							
Infinitive	gapen	zetten	tikken	lachen	blaffen	eisen	
Stem	gaap	zet	tik	lach	blaf	eis	
Past tense (spoken)	/gaptə/	/zetə/	/tɪktə/	/lɑχtə/	/blɑftə/	/Ɛistə/	
Past tense (written)	gaapte	zette	tikte	lachte	blafte	eiste	
Affected by devoicing?							
Spoken	No	No	No	No	No	No	
Written	No	No	No	No	No	No	
Ambiguity in past tense formation based on stem?							
Spoken	Yes [b]	Yes [d]	No[1]	Yes [ɣ]	Yes [v]	Yes [z]	
Written	No	No	No	No	Yes (v)	Yes (z)	
Number of items in *Language Sea*	23	na[2]	48	0	3	27	
Number of incorrect realisations available for error analysis	15,379	na[2]	32,680	0	802	15,888	
Category in study	Semi-consistent					Inconsistent	

[1] /g/only occurs in loanwords in Dutch (e.g., *golf, blog*)
[2]/t/items are excluded in the analyses as these data are not informative for our hypotheses.

Appendix 2

Language Sea past tense items, their token frequency, voicing probability, graphotactic frequency, and number of errors

Type	Infinitive	Correct past tense	Translation	CELEX token frequency	Voicing probability	Voicing consistency	Graphotactic frequency	Number of errors	Nr of types of errors
	Verbs taking -te								
Fricatives	Bruisen	Bruiste	Sparkled	29	0.755	0.245	0.89	190	6
	Eisen	Eiste	Demanded	522	0.755	0.245	0.89	332	8
	Dansen	Danste	Danced	478	0.357	0.643	0.89	540	6
	Dansen	Danste	Danced	478	0.357	0.643	0.89	311	9
	Polsen	Polste	Probed	16	0.357	0.643	0.89	326	11
	Vervalsen	Vervalsten	Forged	3	0.357	0.643	0.89	206	6
	Aanpassen	Pasten aan	Adapted/ changed	296	0.135	0.865	0.89	219	9
	Blussen	Blusten	Extinguished	2	0.135	0.865	0.89	250	9
	Kussen	Kuste	Kissed	1293	0.135	0.865	0.89	1484	22
	Morsen	Morste	Spilled	52	0.135	0.865	0.89	464	9
	Oplossen	Loste op	Solved	161	0.135	0.865	0.89	392	17
	Opvissen	Visten op	Retrieved	39	0.135	0.865	0.89	223	10
	Schorsen	Schorste	Suspended	11	0.135	0.865	0.89	440	13
	Sissen	Siste	Hissed	357	0.135	0.865	0.89	766	15
	Slissen	Sliste	Lisped	9	0.135	0.865	0.89	349	11
	Vergissen	Vergiste	Erred	202	0.135	0.865	0.89	399	11
	Verrassen	Verrasten	Surprised	10	0.135	0.865	0.89	156	15
	Verrassen	Verraste	Surprised	212	0.135	0.865	0.89	234	14
	Vissen	Viste	Fished	130	0.135	0.865	0.89	650	13
	Wassen	Wasten	Washed	52	0.135	0.865	0.89	371	11
	Afstoffen	Stoften af	Dusted	3	0.081	0.919	0.15	237	9
	Beseffen	Besefte	Realized	1516	0.081	0.919	0.15	257	12
	Bluffen	Bluften	Bluffed/ boasted	2	0.081	0.919	0.15	308	10
	Boksen	Boksten	Boxed	3	0.019	0.981	0.89	1021	19
	Botsen	Botste	Collided	153	0.019	0.981	0.89	1278	28
	Kwetsen	Kwetste	Hurt	33	0.019	0.981	0.89	1189	30
	Poetsen	Poetste	Polished	92	0.019	0.981	0.89	1260	27
	Schaatsen	Schaatste	Skated	9	0.019	0.981	0.89	1287	32
	Schetsen	Schetste	Sketched	81	0.019	0.981	0.89	1225	36
Plosives	Aankloppen	Klopten aan	Knocked	145	0.135	0.865	1	349	18
	Betrappen	Betrapten	Caught red-handed	8	0.135	0.865	1	432	12
	Dempen	Dempten	Muffled	10	0.135	0.865	1	607	22
	Kampen	Kampte	Camped	17	0.135	0.865	1	418	23
	Scheppen	Schepte	Ladled/ scooped	184	0.135	0.865	1	1330	37
	Schoppen	Schopte	Kicked	332	0.135	0.865	1	1272	32
	Shoppen	Shopten	Shopped	0	0.135	0.865	1	637	22
	snappen	Snapte	Understood	159	0.135	0.865	1	1314	24
	Soppen	Sopten	Soaked	6	0.135	0.865	1	732	20
	Stoppen	Stopten	Stopped	283	0.135	0.865	1	952	17
	Trappen	Trapte	Kicked	369	0.135	0.865	1	1088	24
	Trappen	Trapten	Kicked	91	0.135	0.865	1	796	15
	Uitstappen	Stapten uit	Got off	616	0.135	0.865	1	405	13
	Verstoppen	Verstopte	Hid	212	0.135	0.865	1	1156	23
	Shoppen	Shopte	shopped	0	0.135	0.865	1	408	22
	Afmaken	Maakte af	Finished	14511	0	1	1	319	11
	Bedanken	Bedankten	Thanked	38	0	1	1	618	21
	Bedekken	Bedekte	Covered	595	0	1	1	747	26
	Benadrukken	Benadrukten	Emphasised	37	0	1	1	420	19
	Bewaken	Bewaakten	Guarded	49	0	1	1	210	9

(*Continued*)

(Continued).

Type	Infinitive	Correct past tense	Translation	CELEX token frequency	Voicing probability	Voicing consistency	Graphotactic frequency	Number of errors	Nr of types of errors
	Bewaken	Bewaakten	Guarded	49	0	1	1	767	18
	Bewaken	Bewaakten	Guarded	49	0	1	1	718	19
	Bewerken	Bewerkten	Edited	35	0	1	1	518	19
	Blaken	Blaakten	Glowed	4	0	1	1	304	13
	Boeken	Boekten	Booked	47	0	1	1	759	10
	Bonken	Bonkten	Banged	43	0	1	1	919	10
	Buitmaken	Maakten buit	Plundered	3422	0	1	1	230	13
	Bukken	Bukten	Crouched	10	0	1	1	1488	23
	Dichtknopen	Knoopte dicht	Buttoned	307	0	1	1	251	14
	Doorstrepen	Streepten door	Crossed through	0	0	1	1	187	13
	Drukken	Drukten	Pushed	269	0	1	1	1110	17
	Harken	Harkten	Raked	4	0	1	1	853	21
	Hinniken	Hinnikten	Neighed	12	0	1	1	380	20
	Hopen	Hoopte	Hoped	1358	0	1	1	399	7
	Klaarmaken	Maakte klaar	Prepared	14511	0	1	1	282	17
	Klikken	Klikten	Clicked	32	0	1	1	816	19
	Kraken	Kraakten	Cracked	107	0	1	1	692	18
	Kraken	Kraakte	Cracked	262	0	1	1	431	12
	Kwijtraken	Raakten kwijt	Lost	842	0	1	1	250	8
	Merken	Merkten	Noticed	343	0	1	1	826	20
	Merken	Merkte	Noticed	3417	0	1	1	996	19
	Mislukken	Mislukten	Failed	49	0	1	1	745	16
	Ontdekken	Ontdekte	Dicsovered	1602	0	1	1	212	8
	Opstropen	Stroopte op	Rolled up (sleeves)	72	0	1	1	448	17
	Opvrolijken	Vrolijkte op	Cheered up	0	0	1	1	189	10
	Playbacken	Playbackten	Lip-synced	0	0	1	1	675	36
	Plukken	Plukten	Picked	40	0	1	1	972	16
	Plukken	Plukten	Picked	40	0	1	1	1273	22
	Prikken	Prikten	Pricked	66	0	1	1	842	21
	Pronken	Pronkte	Boasted	29	0	1	1	512	20
	Rapen	Raapten	Collected	19	0	1	1	781	17
	Reiken	Reikten	Reached	102	0	1	1	510	16
	Schaken	Schaakte	Played chess	17	0	1	1	1011	30
	Schrokken	Schrokten	Wolfed down	3	0	1	1	493	22
	Slopen	Sloopten	Demolished	6	0	1	1	294	13
	Smaken	Smaakte	Tasted	269	0	1	1	472	15
	Snurken	Snurkte	Snoared	54	0	1	1	567	14
	Spieken	Spiekte	Cheated	0	0	1	1	1208	23
	Tikken	Tikte	Ticked	739	0	1	1	1343	26
	Tolken	Tolkte	Translated	0	0	1	1	558	22
	Typen	Typten	Typed	4	0	1	1	574	12
	Uitpakken	Pakte uit	Unpacked	3295	0	1	1	515	20
	Vastmaken	Maakten vast	Tied	3422	0	1	1	213	8
	Verdiepen	Verdiepten	Deepened	31	0	1	1	549	15
	Verslikken	Verslikte	Choked	76	0	1	1	1145	29
	Verzwikken	Verzwikte	Sprained	12	0	1	1	397	22
	Wekken	Wekte	Woke	538	0	1	1	956	34
	Wekken	Wekten	Woke	124	0	1	1	442	16
	Werken	Werkte	Worked	2859	0	1	1	1387	24
	Zakken	Zakte	Failed	877	0	1	1	1217	24
	Playbacken	Playbackte	Lip-synced	0	0	1	1	173	14
	Verbs taking-*de*								
Fricatives	Bonzen	Bonsde	Pounded	207	0.357	0.357	0.11	362	10
	Bonzen	Bonsden	Pounded	25	0.357	0.357	0.11	353	9
	Blozen	Bloosde	Blushed	278	0.755	0.755	0.11	450	18
	Grazen	Graasde	Grazed	19	0.755	0.755	0.11	384	15
	Grijnzen	Grijnsde	Grimaced	777	0.755	0.755	0.11	307	15

(*Continued*)

(Continued).

Type	Infinitive	Correct past tense	Translation	CELEX token frequency	Voicing probability	Voicing consistency	Graphotactic frequency	Number of errors	Nr of types of errors
	Kneuzen	Kneusde	Bruised	6	0.755	0.755	0.11	439	15
	Reizen	Reisden	Travelled	109	0.755	0.755	0.11	278	11
	Verbazen	Verbaasden	Surprised	51	0.755	0.755	0.11	279	16
	Vrezen	Vreesde	Feared	604	0.755	0.755	0.11	408	15
	Vrezen	Vreesden	Feared	166	0.755	0.755	0.11	253	13
	Kerven	Kerfden	Engraved	7	0.775	0.775	0.85	318	12
	Verven	Verfde	Painted	32	0.775	0.775	0.85	586	20
	Klieven	Kliefde	Cleaved	16	0.778	0.778	0.85	198	13
	Proeven	Proefde	Tasted	255	0.778	0.778	0.85	175	5
	Proeven	Proefden	Tasted	35	0.778	0.778	0.85	368	10
	Beleven	Beleefde	Experienced	620	0.992	0.992	0.85	900	34
	Beleven	Beleefden	Experienced	126	0.992	0.992	0.85	378	18
	Beloven	Beloofden	Promised	101	0.992	0.992	0.85	221	5
	Uitdagen	Daagde uit	Challenged	88	0.992	0.992	1	321	15
	Leven	Leefden	Lived	832	0.992	0.992	0.85	577	14
	Overleven	Overleefden	Survived	56	0.992	0.992	0.85	506	16
	Streven	Streefde	Strove	137	0.992	0.992	0.85	239	19
	Volgen	Volgde	Followed	2903	0.953	0.953	1	1171	28
	Walgen	Walgde	Loathed	73	0.953	0.953	1	713	25
	Wijzigen	Wijzigde	Changed	47	0.953	0.953	1	397	24
	Beven	Beefde	Tremored /shook	331	0.992	0.992	0.85	500	24
	Beloven	Beloofde	Promised	1000	0.992	0.992	0.85	1094	31
	Beroven	Beroofden	Robbed	20	0.992	0.992	0.85	184	11
	Doven	Doofden	Extinguished	55	0.992	0.992	0.85	465	18
	Durven	Durfde	Dared	1827	0.992	0.992	0.85	517	17
	Durven	Durfden	Dared	330	0.992	0.992	0.85	527	14
	Erven	Erfde	Inherited	36	0.992	0.992	0.85	577	17
	Geloven	Geloofden	Believed	385	0.992	0.992	0.85	433	12
	Klagen	Klaagden	Complained	88	0.992	0.992	1	224	9
	Kleven	Kleefde	Stuck	136	0.992	0.992	0.85	649	21
	Knagen	Knaagde	Gnawed	71	0.992	0.992	1	1123	24
	Nuttigen	Nuttigden	Consumed	14	0.992	0.992	1	329	15
	Ogen	Oogde	Looked	13	0.992	0.992	1	239	13
	Overtuigen	Overtuigde	Convinced	167	0.992	0.992	1	545	19
	Streven	Streefden	Aimed	81	0.992	0.992	0.85	288	19
	Wuiven	Wuifde	Waved	495	0.992	0.992	0.85	398	17
Glides	Bemoeien	Bemoeiden	Interfered	58	1	1	1	793	20
	Boeien	Boeiden	interested	30	1	1	1	481	7
	Breien	Breide	Knitted	45	1	1	1	493	15
	Opgroeien	Groeiden op	Grew up	334	1	1	1	349	14
	Miauwen	Miauwde	Meowed	16	1	1	1	1008	28
	Naaien	Naaiden	Sewed	14	1	1	1	1230	17
	Roeien	Roeide	Rowed	79	1	1	1	1604	34
	Showen	Showde	Showed	4	1	1	1	949	32
	Sjouwen	Sjouwden	Lugged/ carried	37	1	1	1	716	11
	Snoeien	Snoeide	Pruned	8	1	1	1	1653	38
	Trouwen	Trouwden	Married	137	1	1	1	951	22
	Zwaaien	Zwaaide	Waved	983	1	1	1	1368	31
Liquids	Besparen	Bespaarden	Saved	3	1	1	1	610	17
	Bestellen	Bestelden	Ordered	90	1	1	1	810	16
	Beweren	Beweerden	Claimed	208	1	1	1	384	19
	Boeren	Boerden	Burped	7	1	1	1	1152	23
	Citeren	Citeerde	Cited	152	1	1	1	470	22
	Controleren	Controleerden	Checked	67	1	1	1	272	9
	Delen	Deelde	Shared	782	1	1	1	1290	24
	Delegeren	Delegeerden	Delegated	0	1	1	1	259	14
	Demonstreren	Demonstreerden	protested	21	1	1	1	53	5
	Dwarrelen	Dwarrelden	Floated	62	1	1	1	421	13

(*Continued*)

(Continued).

Type	Infinitive	Correct past tense	Translation	CELEX token frequency	Voicing probability	Voicing consistency	Graphotactic frequency	Number of errors	Nr of types of errors
	Fluisteren	Fluisterden	Whispered	117	1	1	1	513	20
	Gillen	Gilden	Screamed	154	1	1	1	1308	19
	Gniffelen	Gniffelden	Chuckled	2	1	1	1	380	8
	Halen	Haalden	Fetched	564	1	1	1	448	13
	Hameren	Hamerden	Hammered	27	1	1	1	515	17
	Identificeren	Identificeerden	Identified	25	1	1	1	102	9
	Informeren	Informeerden	Informed	29	1	1	1	279	9
	Inspecteren	inspecteerde	Inspected/ checked	130	1	1	1	742	30
	Interesseren	interesseerden	Interested	93	1	1	1	188	14
	Irriteren	irriteerden	Irritated	11	1	1	1	221	12
	Kalmeren	kalmeerden	Calmed down	11	1	1	1	332	9
	Kibbelen	kibbelden	Quibbled	7	1	1		572	17
	Knikkeren	knikkerden	Played with marbles	0	1	1	1	356	15
	Knorren	Knorden	Grunted	10	1	1	1	1028	25
	Afkoelen	Koelde af	Cooled down	32	1	1	1	431	16
	Afkoelen	Koelden af	Cooled down	6	1	1	1	336	10
	Afleveren	Leverde af	Delivered	735	1	1	1	249	15
	Logeren	Logeerden	Stayed over	75	1	1	1	411	6
	Luisteren	Luisterden	Listened	426	1	1	1	825	19
	Naderen	Naderden	Approached	260	1	1	1	502	17
	Ontcijferen	Ontcijferde	Deciphered	12	1	1	1	698	31
	Ontregelen	Ontregelde	Disrupted	20	1	1	1	561	23
	Ontroeren	Ontroerde	Moved/ affected	131	1	1	1	344	16
	Organiseren	Organiseerde	Organised	183	1	1	1	84	7
	Realiseren	Realiseerde	Realized	456	1	1	1	459	18
	Renoveren	Renoveerden	Renovated	0	1	1	1	188	8
	Riskeren	Riskeerden	Risked	9	1	1	1	379	15
	Roddelen	Roddelden	Gossiped	8	1	1	1	375	9
	Rollen	Rolden	Rolled	396	1	1	1	801	10
	Ruilen	Ruilde	Swapped	33	1	1	1	1241	27
	Ruilen	Ruilden	Swapped	43	1	1	1	661	8
	Scoren	Scoorden	Scored	10	1	1	1	143	9
	Scrabbelen	Scrabbelde	Played Scrabble	0	1	1	1	244	15
	Sleutelen	Sleutelde	Tinkered	7	1	1	1	1039	28
	Smeren	Smeerde	Pasted	113	1	1	1	1350	23
	Smeulen	Smeulde	Smouldered	32	1	1	1	1079	41
	Snuffelen	Snuffelden	Snuffled	17	1	1	1	284	6
	Sparen	Spaarde	Saved	77	1	1	1	1150	27
	Spelen	Speelden	Played	893	1	1	1	672	10
	Staren	Staarde	Stared	2213	1	1	1	1091	29
	Stralen	Straalde	Radiated	432	1	1	1	1147	25
	Afstuderen	Studeerden af	Graduated	60	1	1	1	172	7
	Terugsturen	Stuurde terug	Sent back/ returned	987	1	1	1	248	14
	Tillen	Tilde	Lifted	687	1	1	1	368	10
	Volleyballen	Volleybalde	Played volleybal	0	1	1	1	136	12
	Wandelen	Wandelde	Walked/ strolled	531	1	1	1	1137	32
	Weren	Weerde	Excluded/ preserved	126	1	1	1	406	22
	Wellen	Welden	Welled	51	1	1	1	593	29
	Zeulen	Zeulde	Lugged	30	1	1	1	470	19

(*Continued*)

(Continued).

Type	Infinitive	Correct past tense	Translation	CELEX token frequency	Voicing probability	Voicing consistency	Graphotactic frequency	Number of errors	Nr of types of errors
Nasals	Beamen	Beaamden	Confirmed	7	1	1	1	193	7
	Boenen	Boende	Polished/ rubbed	20	1	1	0.99	1159	30
	Gunnen	Gunde	Granted	228	1	1	0.99	1290	26
	Kennen	Kenden	Knew	1026	1	1	0.99	861	16
	Lenen	Leende	Borrowed	172	1	1	0.99	1148	22
	Menen	Meende	Deemed/ believed	2515	1	1	0.99	478	16
	Ontruimen	Ontruimde	Evacuated	20	1	1	1	606	24
	Plannen	Plande	Planned	13	1	1	0.99	386	15
	Rennen	Rende	Ran	1714	1	1	0.99	1374	24
	Opruimen	Ruimde op	Cleaned up	89	1	1	1	462	18
	Schamen	Schaamde	Was ashamed	426	1	1	1	1235	23
	Spammen	Spamde	Spammed	0	1	1	1	711	32
	Spammen	Spamde	Spammed	0	1	1	1	647	27
	Stormen	Stormde	Stormed	268	1	1	1	1555	56
	Verwennen	Verwende	Spoiled	131	1	1	0.99	1238	29
	Bijwonen	Woonden bij	Attended	997	1	1	0.99	221	9