MEASURING EPISTEMIC BELIEFS ABOUT GRAMMAR

ASTRID WIJNANDS[1, 3]*, JIMMY VAN RIJT[2] AND PETER-ARNO COPPEN[3]

[1] University of Applied Sciences, Utrecht, the Netherlands
[2] Tilburg Center of the Learning Sciences, Tilburg University, Tilburg, the Netherland
[3] Radboud Teachers Academy, Radboud University, Nijmegen, the Netherlands

Abstract

In current educational reform much attention is paid to the development of awareness, reflective thinking, and higher order thinking. In language education, the importance of reflectivity for stimulating linguistic awareness and higher order thinking has been emphasized. Crucial for reflective thinking are the underlying assumptions individuals have regarding the nature of knowledge and knowing, so-called epistemic beliefs. Measuring these beliefs is problematic in the domain of language education, where little is known about students' epistemic beliefs about grammar. This study aims to develop an instrument for measuring students' epistemic beliefs in the L1 grammar domain. We therefore transposed a questionnaire for measuring beliefs about history into the domain of grammar. Seven linguistic experts and 300 pre-university students from the Netherlands and Belgium completed this questionnaire about grammar. Exploratory factor analysis extracted two factors, which we argue reflect convergent and divergent thinking. This study shows that students' scores on convergent thinking are higher than the experts' scores, and that students' scores on divergent thinking are lower than the experts' scores. The results also show that students' scores on epistemic beliefs on divergent thinking are higher than their scores on epistemic beliefs on convergent thinking. Possible explanations for these observations are discussed.

Keywords: L1 grammar, epistemic beliefs, reflective thinking, convergent thinking, divergent thinking, pre-university students

1. INTRODUCTION

In recent decades, much research has been undertaken on the development of epistemic beliefs in several domains, such as history (see Maggioni, 2010; Maggioni et al., 2004; Maggioni et al., 2009; Stoel, 2017; Stoel et al., 2017; VanSledright & Maggioni, 2016), mathematics (see Depaepe et al., 2016; Muis, 2004), and science (see Elby et al., 2016). Epistemic beliefs, which are commonly described as beliefs about the nature of knowledge and the nature of knowing are also important in research on learning in general (Hofer, 2002; Hofer & Pintrich, 1997; Muis, 2004, 2007, Muis et al., 2006).
Part of this research is devoted to *measuring* epistemic beliefs. Various methods are used to this end (Mason, 2016), ranging from qualitative interviews (e.g., King & Kitchener, 1994) to paper-and-pencil instruments (e.g., Kuhn et al., 2000) and questionnaires (e.g., Maggioni, et al., 2004; Stoel et al., 2017; Van Rijt et al., 2019).

Stoel et al. (2017) developed a questionnaire for measuring epistemic beliefs about history of students in the highest track of secondary education, based on Maggioni's (2010) 'Beliefs about History Questionnaire' (BHQ). In this study, we adopt the questionnaire by Stoel et al. (2017) and transpose it to another learning domain, namely the linguistic education domain. In doing this, we attempt to measure students' epistemic beliefs on knowledge of language and knowing language, with an emphasis on the grammar domain (in the sense of the linguistic description of language). The research questions for this study are:

1. Can Stoel's et al. (2017) questionnaire be transposed to the linguistics education domain successfully?
    a. To what extent are the results on the questionnaire by Stoel et al. (2017) for measuring epistemic beliefs in history comparable to the results of the questionnaire for measuring epistemic beliefs in grammar?
    b. Does the questionnaire measure an expected valid developmental distinction in epistemic beliefs in grammar?
2. What are pre-university students' epistemic beliefs on language compared to experts' beliefs?

2. MEASURING EPISTEMIC BELIEFS IN THE LINGUISTIC EDUCATIONAL DOMAIN

2.1 Epistemic beliefs and reflective thinking

In research on epistemic beliefs, the powerful relationship between epistemic beliefs and reflective thinking is often stressed (Hofer, 2002; Hofer & Pintrich, 1997; Muis, 2007; King & Kitchener, 1994, 2002, 2004; Maggioni 2010). This relationship works two ways. On the one hand, epistemic beliefs influence the way reflective thinking develops. On the other hand, reflective thinking is the driving force for epistemic belief development (King & Kitchener, 1994, 2004).

The Reflective Judgment Model by King and Kitchener (1994, 2002, 2004) aims to describe the development of reflective thinking and reasoning in a multi-stage model, using core elements of epistemic beliefs: the nature of knowledge (is knowledge perceived as an absolute, best, or relative truth?) and the justification of knowledge (is knowledge justified by authorities, logic or context?). Three main stages are distinguished: (1) a pre-reflective stage, characterized by an epistemic belief that knowledge is absolute and justified by authority and perception; (2) a quasi-reflective stage, where the epistemic belief is, that knowledge is quasi-absolute (idiosyncratic, but solvable) and (also) justified by argumentation; and (3) a reflective stage, where knowledge is seen as a relative truth, the best choice in a given context, and justified by contextual considerations. For reflective thinking this means that pre-reflective thinkers typically will not be able to evaluate from multiple perspectives, and quasi-reflective thinkers will have troubles evaluating conflicting evidence.

Ill-structured problems and divergent information trigger epistemic doubt, which is necessary for development (Bendixen & Rule, 2004; Kienhues et al., 2016; King & Kitchener, 1994). Since no clear-cut solutions to these problems are given, students are bound to challenge their epistemic beliefs, and they are likely to change their usual ways of thinking and adapt their beliefs.

In current educational reform, concepts like awareness, reflective thinking and higher order thinking skills are gaining momentum in general. For the language education domain, this is asserted in a broad context by Boivin et al. (2018), and it is further developed for specific language domains: for the Anglophone situation: ACARA (2009), Denham (2020), Denham & Lobeck (2010), Derewianka (2012), Exley & Mills (2012), Trousdale (2010); for the Francophone situation: Boivin (2018); for the Hispanic situation: Camps &

Fontich (2019); Fontich & García-Folgado (2018); for the Dutch situation: Curriculum.nu (2019); Meesterschapsteams (2018), Van den Broek (2020); Van Rijt (2020), VVKSO (2014). Therefore, the relationship between reflective thinking and epistemic beliefs is becoming more and more important. From current research it seems that high school students generally exhibit pre-reflective thinking skills (see King & Kitchener, 2004), whereas most college students show quasi-reflective thinking skills (King, 2009; King & Kitchener, 2002; Mierewald et al., 2017). In order to develop reflective thinking for pre-university school students, attention should be paid to the development of epistemic beliefs. Measuring these beliefs is a prerequisite for this. To our knowledge, no research has been conducted on measuring epistemic beliefs about grammar.

2.2 Measuring epistemic beliefs about grammar

Stoel's et al. (2017) questionnaire on historical epistemic beliefs seems to offer a good basis for constructing a questionnaire for assessing epistemic beliefs for the domain of grammar. According to Muis et al. (2006), domains are comparable if they are alike in the nature of the problems involved (ill-structured or well-structured), the body of the theory (paradigmatic/hard or non-paradigmatic/soft), and the way the theory progresses (theoretic development/pure or practical/applied). Muis et al. (2006) characterize the history domain as ill-structured, soft, and pure. The grammar domain can be characterized in the same way: although decontextualized grammar (parsing) exercises suggest that grammar problems are well-structured and that reasoning is paradigmatic/hard, the problem of grammatical choice is an ill-structured problem, and even for parsing exercises, research into expert linguistic analysis (Dielemans & Coppen, 2021) shows that linguistic experts use both convergent and divergent reasoning to solve analytical problems. This makes linguistic reasoning about grammar also non-paradigmatic/soft. Finally, prescriptive grammar knowledge may appear to be governed by practical considerations, but expert grammar knowledge is all about theoretical development, just as in the history domain.

Stoel's et al. questionnaire is also a good basis for our purposes because it targets students in the highest track of secondary education, which is also our focus, and it is based on theories about epistemic beliefs in relation to reflective thinking.

3. METHOD

3.1 Participants

For this study we invited seven linguistic experts to fill in the questionnaire, firstly in order to replicate Stoel's et al. (2017) study, and secondly to be able to compare secondary school student scores with accomplished linguist scores. These linguistic experts were full professors of Dutch or English linguistics or Second Language Acquisition. Two professors were emeritus. The participating professors gave active consent to use their data anonymously in the research.

Five Dutch teachers from four different secondary schools and six Belgian teachers from three different secondary schools voluntarily participated in this study. All teachers were grade 1 certificated, which means that they can teach in all levels of secondary education. Eight teachers were university graduates, and the rest of three graduated from universities of applied sciences. Their experience as teachers in Dutch language and literature ranged from six to thirty years. The study of this article is part of a larger study on the development of reflective thinking about language issues (see also Coppen et al., 2019; Wijnands et al., 2021). In this larger study, three Belgian teachers (from two different schools) participated in a Professional Learning Community (PLC) and three Dutch teachers (from three different schools) participated in another PLC. The remaining five teachers were approached separately by the first author to participate in this part of the study.

The questionnaire was reviewed beforehand with two Belgian teachers from the PLC (see also Section 3.3). All teachers received information about the objectives of this study and were asked to follow a protocol when instructing their students (see Appendix A and Section 3.2).

A total of 302 pre-university students participated in this study. The students were asked to sign a consent form, in which it was stated that their data would be used anonymously for scientific research. In total, 300 students signed this consent. In case of students younger than 16 years, their parents signed the consent. Two Belgian students withheld their consent and were consequently left out the data collection. The 300 participating students came from two Dutch-speaking areas: the Netherlands and Belgium (Flanders), see Table 1.

In the Dutch curriculum, grammar is taught in the 7th to 9th grade (12-15 years old). In Belgium, students receive grammar education from the 7th to 12th grade (12-18 years old). In both countries grammar teaching is mostly

traditional, focusing on parsing sentences and labelling parts of speech, using a traditional Latin-based body of grammar knowledge (cf. Van Rijt, 2020). The present study was approved by the ethical committee of the Radboud University in Nijmegen, under reference number 18U.019822.

Table 1. Participants

	Belgian students	Dutch students		
	11th grade	11th grade	12th grade	total
	16 – 17 years	16 – 17 years	17 – 18 years	
Test moment 1	50	41	53	144
Test moment 2	78	78	-	156
total	128	119	53	300

3.2 Procedure

The experts completed the questionnaire in Qualtrics (an online survey platform) without time limit. For the students, there were two test moments (see Table 1). In the first test moment, 144 students filled in the questionnaire with paper and pencil. In the second test moment, 156 students completed the questionnaire in Qualtrics. The choice of two ways of testing arose from the preference of the teachers. In the first testing moment the preference was for paper and pencil, in the second testing moment for Qualtrics. In both sub-groups the teachers of the students were asked to follow a protocol (see Appendix A) in which their students were given about 10 minutes for completing the questionnaire.

An independent T-test revealed no significant difference between the test conditions (paper-and-pencil vs Qualtrics). This T-test was computed from the 11th grade student results only, because for the 12th graders, only one condition was applied. Because the results from both test conditions did not differ, we combined them in the remaining analyses. There is also evidence in the literature to suggest that migration from paper-and-pencil to online surveys does not have a substantial effect on the outcomes of the measurement (Barzilai & Weinstock, 2015; Roberts, 2007).

3.3 Development of the questionnaire

Our starting point for developing the questionnaire for the domain of grammar was the concept questionnaire consisting of 26 items developed by Stoel et al. (2017, p. 126). The first author, a teacher educator of the Dutch language, and the third author of this paper, a full professor of linguistics, developed a first transposition from the history domain to the linguistic domain. In this transposition, several standard rephrasings were applied. For example, a phrase such as 'to write adequately about history' was converted to 'analyze words and sentences correctly'. We converted 'historical events' to 'language intuitions', since according to Chomskyan linguistic theory (Chomsky, 1965), language utterances can only be considered as indirect evidence of the language that resides in the human brain; language intuitions, as more direct evidence, come closer to the notion of 'linguistic facts'. Therefore, language intuitions are comparable to historical facts. They have to be explained by theory (for a full overview of the conversions, see Appendix B). Three items (1, 9, and 10) were completely reformulated, while preserving the essential epistemic content. This first transposition was validated by two Flemish teachers to prevent a language variant effect (the Belgian language and context differs from the Dutch variants, and both authors were from the Netherlands). As a result, some words and formulations were adapted. For instance, the word *ontleden* 'parsing' is seldom used in Belgium, so the more general term *analyseren* 'analyze' was chosen.

In order to discover deficiencies in the formulation of the test items, we piloted the questionnaire with two students. One student was from pre-university education (12th grade) and the other student was a third-year bachelor university student. They scored the items in a thinking aloud protocol. On basis of the transcripts of these sessions some items were improved in the formulation (see Appendix C for the final questionnaire). In line with Stoel (2017) and Maggioni et al. (2009), the items had to be scored on a six-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree).

Finally, the authors of this paper individually categorized the 26 items on language into pre-reflective, quasi-reflective, and reflective items. Fleiss' kappa showed that there was an initial moderate agreement between the authors' judgements, $\kappa =.531$ (95% CI, .373 to .690), $p < .001$. Disagreement only occurred between adjacent scales (e.g., between the pre-reflective and quasi-reflective scales) and not between the pre-reflective and reflective scales. Because these three scales represent a developing continuum, disagreement might be considered as borderline cases of a particular scale. The

disagreed items were subsequently resolved through discussion, resulting in a full agreement on the classification.

3.4. Data analysis

We analyzed the data in different steps. The first step concerned the experts' scores in order to find out whether the adjusted items test the intended reflective attitudes. We followed Maggioni et al. (2009) and Stoel et al. (2017) in not applying extensive statistical analyses on the scores of the seven experts, because of the relatively small sample size. We calculated the average scores on the three scales we conceived beforehand. In order to assess the reliability of these scales we calculated Cronbach's alpha for experts' scores.

The second step concerned the analysis of the students' scores on the questionnaire. We conducted an exploratory factor analysis (Principal Axis Factoring) with Varimax rotation and Kaiser Normalization. Because we realize that our data are nested since seven schools were involved, we also conducted a Multilevel Confirmatory Factor Analysis (MCFA) using R (Huang, 2017). This MCFA did not result in a positive outcome, which is probably due to the small size of the nested structure. In order to find out whether the factors found in the exploratory factor analysis were valid for the different schools, we conducted factor analyses (Principal Axis Factoring) with Varimax rotation separately for those schools. The results of these separate analyses were to a large extent comparable with the integral analysis. From this we conclude that the integral analysis has sufficient validity.

The third step in our data analysis concerned the comparison (using independent T-tests) of the mean scores for the Belgian and Dutch students, since both the content and the educational practice in the curricula differ (Bonset & Hoogeveen, 2010; Knop & Van Laere, 2017; VVKSO, 2014).

In the third step we investigated the differences between the scoring by students from the 11th grade and 12th grade, because it seems likely that reflectivity develops over time. In the final step we compared the mean scores of the experts (n = 7) with the mean scores of the students in order to find out to what extent experts' scores differed from students' scores.

4. RESULTS

4.1 Experts' scores

In order to validate the stability of the adjusted items of the questionnaire, we asked linguistic experts to score the items. Because of their scientific attitude towards language, we expected the linguistic experts to score high on agreement with the reflective items, lower on quasi-reflective items, and much lower on pre-reflective items.

Figure 1 shows that the seven linguistic experts indeed score high on reflective items and low on pre-reflective items. Although there is more variance between the experts' scores in the pre-reflective (M = 2.26, SD = 1.15) and quasi-reflective items (M = 3.06; SD = 1.51), the scores are considerably lower than the scores on the reflective items (M = 4.77; SD = 1.01) (see Appendix C for the raw scores). Reliability analysis showed that Cronbach's alpha was excellent for the pre-reflective (.95) and reflective (.95) scales and good for the quasi-reflective (.82) scales.

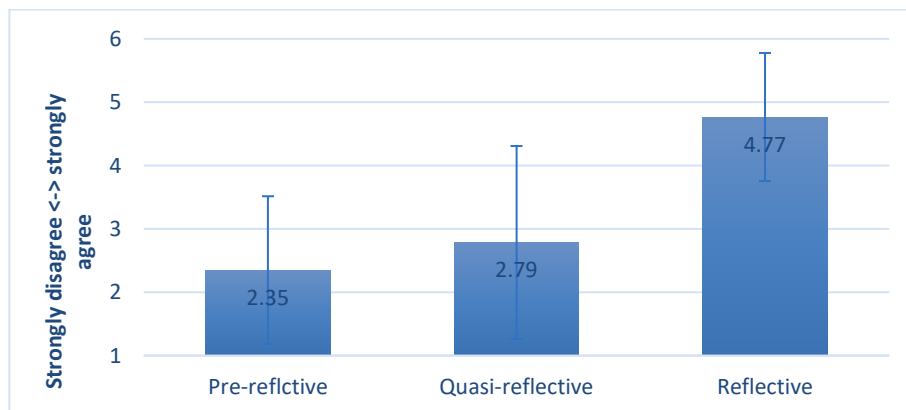*Figure 1: Experts' average scores on pre-reflective, quasi-reflective and reflective items (n = 7).*

4.2 Students' scores

The 26 items of the questionnaire (n = 300) were factor analyzed using Principal Axis Factoring) with Varimax rotation. After removing items with a loading below .3, two factors with eigenvalues over one emerged, explaining a total of 36.79% of the variance, see Table 2.

Table 2. The two factors obtained in the exploratory factor analysis (n = 300)

		Factor 1	Factor 2
20.	A word or sentence is analyzed correctly when you have gone through all the steps. (P)	.622	
5.	*Analyzing words and sentences is a good way to develop inquiry skills. (R)*	.497	
21.	When a language rule is written in your textbook, you can be nearly certain that it is true. (P)	.451	
25.	**In a thorough analysis all the causes of your own language intuitions will become clear. (Q)**	.440	
1.	The goal of analyzing words and sentences is to find the correct solution. (P)	.346	
22.	*The analysis of words and sentences is a critical search for form and meaning. (R)*		.504
15.	*A good analysis of words and sentences must take different perspectives on form and meaning into account (What do you think? What do other people think? What are the rules?) (R)*		.446
26.	*In many cases, linguistic experts will continue to discuss language rules and language intuitions. (R)*		.355
4.	*When analyzing words and sentences you must learn to deal with conflicting information. (R)*		.342
24.	*Even when consulting the same sources, linguistic experts often arrive at different analyses. (R)*		.330
	Eigenvalues	2.13	1.55
	% of variance	21.26	15.53

Note 1: The items were tested in Dutch. The English translation is as close as possible to the original items. However, due to translation issues, there might be some differences in interpretation. Note 2: P = Pre-reflective item; Q = Quasi-reflective item; R = Reflective item. Note 3: roman-red = pre-reflective item; bold-orange = quasi-reflective item; italic-green = reflective item.

The factor structure by exploratory analysis was different from the original structure of the questionnaire. Factor 1 was comprised of five items reported on a six-point Likert scale with factor loadings from .346 to .622 (see Table 2). This factor consists of 3 pre-reflective items (displayed in red, roman), one quasi-reflective item (displayed in orange, bold), and one reflective item (displayed in green, italic). Factor 2 consisted of five reflective items (displayed in green, italic) with loading factors from .330 to .504 (see Table 2).

Internal consistency for the two scales was examined using Cronbach's alpha. The reliability of the scale for Factor 1 was .57, and for Factor 2 was .48, which is commonly interpreted as poor (Factor 1) or even unacceptable (Factor 2). A substantial increase could not be achieved in alpha for the scales by eliminating more items. We will return to this in the discussion.

4.3 Differences between students' groups

There was no significant difference in Factor 2 scores between the students of the two different countries (t(245) = 0.984, p = .33). Factor 1 showed a narrow significant difference between the Dutch and Belgium students (t(245) = 1.972, p = 0.050), see Table 3. As expected, 12th grade students attribute less value to items about convergent thinking than 11th grade students (t(170) = 3.481, p = .001), see Table 3. The mean scores by the experts were, also as expected, lower than the 12th grade students for this factor. In Factor 2, divergent thinking, we found no significant difference between the 11th and 12th grade students (t(80) = -0.161, p= .872). However, the experts' mean score was 0.3 higher than the students' mean score. See also Figure 2 for the differences between the students and the experts.

Table 3. Means and standard deviations per sub-group and samples independent T-test per factor per students' sub-group.

		Students					Experts			
		n	M	SD	t	p	M	SD	min	max
	11th grade Belgian students	128	4.18	0.71	1.972	0.050				
Factor 1 *Convergent thinking*	11th grade Dutch students	119	4.02	0.57	3.481	0.001				
	12th grade Dutch students	53	3.65	0.77						
	Total	300	4.02	1.20			3.34	0.32	3.0	3.8
	11th grade Belgian students	128	4.58	0.52	0.984	0.717				
Factor 2 *Divergent thinking*	11th grade Dutch students	119	4.51	0.51	-0.161	0.872				
	12th grade Dutch students	53	4.52	0.66						
	Total	300	4.54	0,98			4.89	0.65	3.6	5.4

4.4 Differences between students and experts

Experts valued items of Factor 1 lower (M = 3.34) and items of Factor 2 higher (M = 4.89) than the students did (see Table 3 and Figure 2). Reliability analysis showed that Cronbach's alpha was good for Factor 1 (.75) and excellent for Factor 2 (.98).
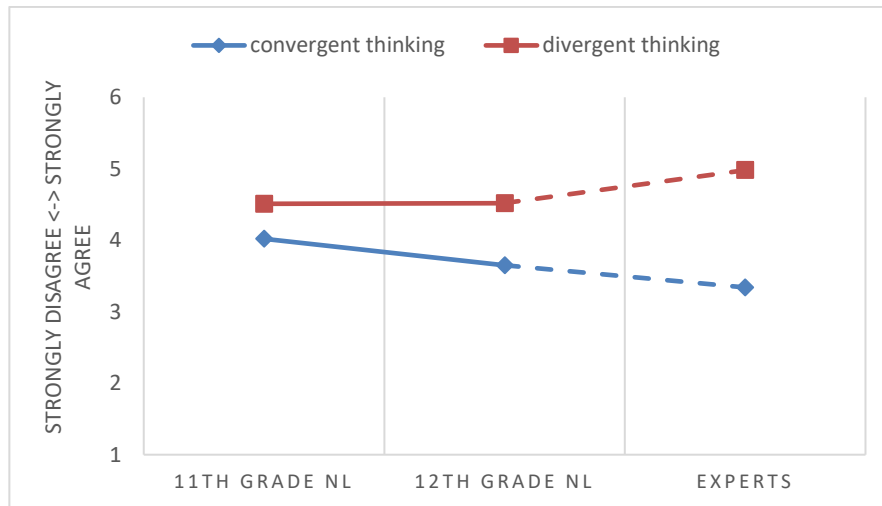
Figure 2: Scores on Factor 1, convergent thinking, and Factor 2, divergent thinking, by students from 11th and 12th grade from the Netherlands, and by linguistic experts. The dashed line represents an uncertain period. Lines do not represent development, but differences between groups in different stages of development.

5. DISCUSSION

The current study aimed to answer two research questions. In the first research question we examined to what extent Stoel's et al. (2017) questionnaire could be transposed successfully to the linguistic education domain. The second question aimed at the results from using the questionnaire to measure pre-university students' epistemic beliefs on grammar.

5.1 Interpretation of the main results

The questionnaire had been constructed for measuring pre-reflective, quasi-reflective, and reflective epistemic beliefs. From the results, it appears that factor analysis did not reveal three but two factors, each consisting of five items.

Factor 1 included three items originally intended as pre-reflective, namely (20), (21) and (1) (see the roman-red items in Table 2). Item 20 having the highest loading strikingly characterizes pre-reflective epistemic beliefs ('A word or sentence is analyzed correctly when you have gone through all the steps'). The statement focuses on a single correct solution, and a fixed and certain method of obtaining that one solution. In addition, two other non-

pre-reflective items load on this factor, namely item (25) and (5). Item (25) ('In a thorough analysis all the causes of your own language intuitions will become clear') was intended to measure quasi-reflective thinking, because it focused on analysis, which arguably would entail some kind of argumentation. However, in hindsight it is explainable why this item loads on this factor. In general, the quasi-reflective stage is a transition stage between pre-reflective and reflective. This entails that there will be boundary cases, and the awareness that some kind of analysis is possible is still in accordance with a pre-reflective mind believing that the correctness of such an analysis is fully determined by some authority. In addition, similar research shows that quasi-reflective items are difficult to measure as a group. From research (e.g., Maggioni, 2010; Mierwald et al., 2017; Stoel et al., 2017) it appears that quasi-reflective items can load together with pre-reflective ones.

Item 5 ('The analysis of words and sentences is a good way to develop your inquiry skills') also loads on the first factor, although it was originally intended to reflect reflective epistemic beliefs. It may be that students are just repeating the words of the teacher or general beliefs here. Grammar education is often said to develop abstract thinking skills (Honda & O'Neil, 2008; Hudson, 1999, 2004; Hulshof, 2014; Van Gelderen, 2010; Van Rijt, 2020), and this may be one of their teacher's standard answers when they ask why they have to do grammar exercises. So they may still be referring to the authority of the teacher, which is typical for a pre-reflective stage.

It seems therefore that Factor 1 is largely compatible with the pre-reflective stage. Factor 2 is even more consistent. All items were originally intended as belonging to the reflective stage. And as with Factor 2, the item with the highest load (Item 22, 'The analysis of words and sentences is a critical search for form and meaning') most typically represents reflective thinking, focusing on critical thinking.

This result begs the question whether the questionnaire in fact measures two kinds of beliefs instead of three. Maggioni (2010) and Mierwald et al. (2017) also revealed only two factors, which they coined naïve and nuanced (loosely similar to pre-reflective/quasi-reflective and reflective). However, looking at the items loading on our two factors, another interpretation emerges. It looks like the items loading on Factor 1 are all in some way associated with convergent thinking, whereas the items loading on Factor 2 are more related to divergent thinking (Cropley, 2006).

This is a somewhat unexpected, but interesting result. In Wijnands et al. (2021) a pedagogical template was based on two pedagogical models, one of which involved cognitive learning, and the other one reflective development.

Wijnands et al. (2021) also interpreted the two dimensions of this model as convergent and divergent. In convergent thinking, when working towards conceptual understanding, the best answer is chosen, and there is no room for alternatives. As Cropely (2006, p. 3) states, '…Answers are either right or wrong'. In divergent thinking, when working towards broader views, multiple answers are possible, which may be novel or even unusual. Divergent answers may differ among individuals, but they can, as Cropley (2006, p. 4) emphasizes, 'be of equal value'. In this sense, the results from the current study are compatible with previous work, both from our own research group and beyond.

5.2 Transposition of the questionnaire

In order to answer the first research question, we formulated the following two sub-questions:
a. To what extent are the results on the questionnaire by Stoel et al. (2017) for measuring epistemic beliefs in history comparable to the results of the questionnaire for measuring epistemic beliefs in grammar?
b. Does the questionnaire measure an expected valid developmental distinction in epistemic beliefs in grammar?

For the first sub-question, we need to look at the main results of the study by Stoel et al. (2017). In their study four main results were found: (a) a factor analysis revealed three factors, labelled as 'objective nature of knowledge', 'objective nature of knowing', and 'historical methodology'; (b) students had positive scores both on 'objective nature of knowing items' and 'historical methodology'; (c) experts valued reflective items higher than pre-reflective items; (d) experts valued reflective items higher and pre-reflective items lower than the students.

In contrast, factor analysis in our study came up with two factors, which we labelled as convergent and divergent thinking. At first glance, this result does not appear to be in line with Stoel et al. (2017). However, one of their factors (nature of knowing – nuanced) seems to correspond with our factor divergent thinking, whereas the two other factors (nature of knowing naïve (objective) and nature of knowledge - objective) all seem to have a convergent character. So there seems to be at least some correspondence between our results. The differences could be due to domain specific properties of the linguistic domain and the history domain. The items in Stoel's et al. (2017) factor nature of knowing – naïve all focus on the reliability of sources. Three

of the corresponding items in our questionnaire (10, 16, 17) did not load on any factor. This could be due to the fact that in the domain of secondary school linguistics, sources are less clear than in the history domain: the fact that one's own intuition can be a source is virtually absent in current grammar teaching practice, and even language reality as a source is not always used, since for example grammar teaching often focuses on prescriptive sources (cf. Berry, 2015; Hudson, 2004).

As with Stoel et al (2017), in our study we found that students have high scores on both convergent and divergent items (convergent M = 4.02; divergent M = 4.54), that experts valued items originally intended as reflective higher than items originally intended as pre-reflective (reflective items M = 4.89; pre-reflective items M = 3.34), and that experts valued the items originally intended as reflective higher than the students (experts M = 4.89; students M = 4.54), whereas they valued the items originally intended as pre-reflective lower (experts M = 3.34; students M = 4.02).

In conclusion, our first sub-question can be answered affirmatively. We can conclude that the results on our questionnaire for measuring epistemic beliefs in grammar were comparable to the questionnaire by Stoel et al. (2017) for measuring epistemic beliefs in history.

For the second sub-question, we examined whether our questionnaire measured an expected valid developmental distinction in epistemic beliefs in grammar. Stoel et al. (2017) found significant differences in epistemic beliefs between students in the highest track of higher general continued education preparing for universities in applied sciences (mean age 17 years old) and in the highest track of pre-university education (mean age 18 years old). The pre-university students scored pre-reflective items lower and reflective items higher than students in higher general continued education. In our study, we investigated differences between students in the 11th and 12th grade, the highest grade, of pre-university education.

Our results indeed suggest a similar development in convergent thinking: Dutch 12th grade students agreed less with items about convergent thinking than 11th grade students. Following Stoel et al. (2017), we assume this is a result of the fact that 12th graders are less focused on finding the right solution or analysis since they have progressed more in their academic development (see also Stoel et al., 2017, p. 128). In addition to this supposed development, we observed (cf. Table 4, Figure 2) that experts show even less affinity with convergent thinking. This corroborates the suggestion that a decreasing affinity with convergent thinking indeed is part of a development towards academic thinking.

With respect to divergent thinking, we did not observe a significant difference between 11th graders and 12th graders, in line with the results from Stoel et al. (2017). However, experts showed a substantially higher score on divergent thinking.

In summary, it seems that our questionnaire is a valid measure for development of convergent and divergent thinking. Results are in line with expectations, and comparable to the study by Stoel et al. (2017). This means that our second sub-question can be answered affirmatively.
Based on the results on the two sub-questions, we can conclude that a transposition of Stoel's et al. (2017) questionnaire is possible for the linguistic education domain. The final questionnaire for measuring epistemic beliefs on language consists of ten items and measures the degree of convergent thinking and divergent thinking.

5.3 Epistemic beliefs on language

Our second research question was: what are pre-university students' epistemic beliefs on language compared to experts' scores? In order to answer this question, we first consider the experts' beliefs on convergent and divergent thinking. From the results (see Table 3), it appears that experts had a low score on convergent thinking and a high score on divergent thinking.[1] This confirms our expectation that experts are trained to consider different perspectives on language issues to make a well-reasoned judgment on how to assess this language issue (see Dielemans & Coppen, 2021). However, we also observe that experts score still rather high on convergent thinking. Hence, apparently, convergent thinking is still an important aspect of expert thinking, notwithstanding with their decreased affinity. This is in line with King and Kitchener's (1994) assertion that reflective thinking is not only a matter of multi-perspectivity (which we can identify as divergent), but also of evaluation (which essentially is convergent, cf. King & Kitchener, 1994, 2002, 2004). Convergent thinking is also acknowledged as an important part of creative thinking by Cropley (2006) and Jaarsveld et al. (2012).

---

[1] *One of the experts pointed out (personal communication) that he had difficulties with filling in the questionnaire because of his interpretation of 'linguistic intuitions'. According to his expertise, the analysis of words and sentences always refers to the linguistic intuitions of mother-tongue speakers. We acknowledged this remark, but this expert was the only one with deviating scoring, especially on the items 1, 15, 24 and 26. When these experts scores would be eliminated, the difference between convergent and divergent thinking would increase.*

Comparing the experts' scores with the students' scores on convergent and divergent thinking, we see that students' scores on convergent thinking are higher than the experts' scores and that students' scores on divergent thinking are lower (see Table 3 and Figure 2). However, we also observed that students' scores on epistemic beliefs on divergent thinking are higher than their scores on epistemic beliefs on convergent thinking.

One explanation for this would be that the scores on divergent thinking are a kind of baseline score, reflecting general (non-domain specific) beliefs on the nature of knowing and knowledge. The fact that there is no observable difference between 11th graders and 12th graders points in this direction. If the baseline is already fairly high, a significant increase might be less probable, since it would require more expert knowledge in linguistics, which is not a part of the secondary school curriculum in the Netherlands (see Van der Aalsvoort, 2016; Van Rijt & Coppen, 2017). The responses to the divergent items might have been influenced by students' ideas about how knowledge is gained, as Stoel et al. (2017) also describe for the historical domain. Students might have interpreted these items as descriptions of critical thinking strategies that lead to a correct or unambiguous answer (see Stoel et al., 2017, p. 131).

The base score on the divergent items could also be explained by students giving socially desirable answers. The divergent items might have sounded like a possible way of studying language and therefore would have received their high score. However, if students answered the items in a socially desirable way, this should not only have been the case for divergent items, but also for convergent items. Students might have been influenced by the school setting in which the focus is on giving the correct answer, resulting in higher scores on convergent items than is currently the case.

Another explanation for the base line scores on divergent thinking might be that the 12th graders are subject to the final examinations, which could discourage divergent thinking and favour convergent thinking. A similar effect can be seen in the study by Yuen Lie Lim (2011) on reflective thinking among university students (16-21 years). Year 3 students were found to be less reflective than year 1 students because they focused more on the requirements for completing a course than on reflective thinking. We leave this matter open for future research.

A final explanation is the one Stoel et al. (2017) assume to account for the difference they found between senior general secondary education (11th grade) and pre-university education (12th grade). Stoel et al. found that pre-

university students scored higher on divergent thinking (nuanced in their terminology), and ascribed this to the higher development in academic thinking. Since we only have scores of pre-university students, it could be that the difference they found was more a difference between senior general secondary education and pre-university education than between 11th grade and 12th grade. Pre-university education would then always score high on divergent thinking.

Our results also show a small significant difference between 11th grade Belgian students and 11th grade Dutch students. Belgian students appeared to value beliefs on convergent thinking higher than Dutch students. An explanation for this might be that in the Dutch curriculum grammar is only taught in the 7th until 9th grade of pre-university education whereas in Belgium grammar is taught throughout the whole secondary school curriculum (so including 11th grade). Belgian students might therefore be more focused on the correct answers than Dutch students because they are still subject to grammar teaching, which is strongly focused on testing (Van Rijt, 2020). A second explanation could be related to the different status of the prescriptive norm of Standard Dutch. In Belgium, Standard Dutch is more of a societal issue in competition with dialectical or supra-regional variants ('tussentaal'). Although supra-regional variants are used by teachers and students in the school environment, Standard Dutch is also seen as the only acceptable language variety at schools (Delarue, 2016). Using Standard Dutch in Belgium is strongly associated with high prestige and good education (Delarue, 2016, Lybaert, 2015). This is more in line with convergent thinking about grammar and pre-reflective attitudes.

We did not find any difference between Dutch and Belgian students in their divergent thinking epistemic beliefs on the analysis of words and sentences. This confirms our aforementioned assumption that the scores on divergent thinking are a kind of baseline score.

To summarize, this study aimed to measure the epistemic beliefs of students in secondary education about grammar by means of a questionnaire. Given the results on the two research questions, it seems that our transposition of the questionnaire by Stoel et al. (2017) offers a valid instrument that reveals a clear difference between secondary school students and experts, and to a lesser extent a difference between different categories of students.

5.4 Limitations and future research

This study is the first in exploring domain-specific epistemic beliefs of pre-university students on grammar. Other studies explored general epistemic beliefs, or epistemic beliefs of teachers (e.g., Van Rijt et al., 2019), and even in combination with grammar tests, but no specific epistemic beliefs of students on grammar (see e.g., Elsner, 2020).

Although our research questions can be answered with a fair degree of confidence, conclusions should still be considered with some caution. Research into epistemic beliefs by means of questionnaires is never without problems (see Maggioni, 2010). For one thing, to further validate the questionnaire, it should be replicated more often. From other questionnaires we know that students' responses are sometimes inaccurate because of the difficulty of the items (see Barzilai & Weinstock, 2015; Mason, 2016). In our questionnaire, some of the items also seemed to cause interpretation problems. More sophisticated (mixed methods) measurements are recommended for further research. For example, interviewing students after completing the questionnaire could provide more information about how they scored the items. A promising instrument for measuring epistemic beliefs is the scenario-based instrument created by Barzilai and Weinstock (2015). In this instrument epistemic beliefs are measured using dilemmas, each with two conflicting expert reactions about which students must then score statements (see also Mason, 2016). However, this research is outside the scope of our study for now.

It was shown in Section 4.1 that for the experts' scores the reliability of the larger scales was good to excellent. However, in Section 4.2 it was noted that the Cronbach's alpha on the students' scores were considered poor. This can be a result of the small number of items on each scale (see Field, 2013, p. 709). However, these low Cronbach's alphas might also be an indication that the students found it difficult to score the items because of their lack of knowledge on this subject, which is not the case for linguistic experts. It is reasonable to assume that the experts fully understood the content of the items, resulting in a larger reliability. This would mean that a better linguistic schooling would also result in more reliable scores for students.

The questionnaire we transposed to the domain of grammar originally consisted of 26 items. In our study, factor analysis only revealed a total of 10 relevant items, equally divided over two factors. In addition to the observation mentioned above that this is mainly caused by boundary cases; a further reason for this small number of items might be that students find it difficult

to score items about 'the analysis of words and sentences' because grammar teaching is often not really implemented as analysis, but rather as a low-level classification exercise on the basis of rules of thumb, without much reasoning. Therefore, students might not even think about grammar as analysis of words and sentences. It would be interesting to explore in future research whether a replacement of phrases as 'analysis of words and sentences' by 'grammar' would lead to an increase in the number of relevant items. If the interpretation of items is a problem for students, it would be expected that interventions in which students are confronted with real linguistic analysis (in contrast to the usual rules of thumb) would also lead to more relevant items.

Another possible limitation is the number of students involved in this study. Our study is based on the responses of 300 students, whereas Stoel et al.'s (2017) study is based on 922 students. However, for factor analysis a minimum of 300 participants is recommended (Field, 2013, p. 684; Tabachnick & Fidell, 2013, p. 618). Therefore, despite the difference in numbers, our study does not seem to be underpowered.

Another limitation is that 12th grade Belgian students were not included. As a result, we cannot compare Belgian 12th grade students to Dutch 12th grade students. However, we would expect that the difference we observed for the 11th graders pertains, since we explained this difference by differences in the national curricula, which remain the same (the status of standard language, the curriculum structure with final exams).

In order to improve our questionnaire, and to put it to use in further research on reflective development, an obvious way to proceed is to use the questionnaire to measure reflective development in design research. To this end, an intervention should be devised to confront students with ill-structured language problems, since these are known to create the experience of epistemic doubt, which is a pre-requisite for epistemic change (cf. Kienhuis et al., 2016, for an example from the philosophy domain).

In Wijnands et al. (2021) a pedagogical template for developing convergent and divergent thinking about ill-structured language problems has been described. In Wijnands et al. (forthcoming), the implementation of this template will be discussed. In this template, students learn how to use primary sources such as their own language intuitions and language reality (for instance corpus data), and secondary sources, such as language advice and reference grammars to investigate an ill-structured problem more deeply. Students can work on assignments in which the focus is on the development of

convergent thinking, meaning that they work towards an analysis of the language problem. Students can also work on assignments in which the focus is on the development of divergent thinking, meaning that they investigate the language problem from different perspectives, using more different sources. When working on these divergent thinking assignments, students will be confronted with epistemic doubt which might lead to epistemic change in their reflection on the knowledge of language and knowing language. Such an epistemic change could then be measured by the questionnaire. This study has established that the instrument is a promising way to do this.

ACKNOWLEDGEMENTS

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

NOTES ON CONTRIBUTORS

Astrid Wijnands is a lecturer of Linguistics and a teacher trainer at the HU University of Applied Sciences as well as a PhD-candidate at Radboud University Nijmegen.

Jimmy van Rijt is an assistant professor of the learning sciences at Tilburg University, whose research is primarily aimed at the pedagogical and conceptual enrichment of L1 grammar teaching.

Peter-Arno Coppen is a full professor of Linguistics at Radboud University Nijmegen with a special interest in grammar education and grammar teaching.

ORCID

Astrid Wijnands https://orcid/org/0000-0002-4988-4789
Jimmy van Rijt https://orcid/org/0000-0003-2665-3707
Peter-Arno Coppen https://orcid/org/0000-0001-5847-5322

REFERENCES

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2009). *Framing English*. ACARA.

Barzilai, S., & Weinstock, M. (2015). Measuring epistemic thinking within and across topics: A scenario-based approach. *Contemporary Educational Psychology 42*, p. 141–158. https://doi.org/10.1016/j.cedpsych.2015.06.006

Bendixen, L.D., & Rule, D. C. (2004). An integrative approach to personal epistemology: A guiding model. *Educational Psychologist, 39*(1), 69-80. https://doi.org/10.1207/s15326985ep3901_7

Berry, R. (2015). Grammar myths. *Language Awareness 24*(1), 15-37. https://doi.org/10.1080/09658416.2013.873803

Boivin, M.-C. (2018). A review of the current empirical research on grammar instruction in the francophone regions. *L1-Educational Studies in Language and Literature 17*, 1-48. https://doi.org/10.17239/L1ESLL-2018.18.04.03

Boivin, M.-C., Fontich, X., Funke, R., García-Folgado, M.-J., & Myhill, D. (2018). Working on grammar at school in L1 education: Empirical research across linguistic regions. *L1-Educational Studies in Language and Literature 18*, 1-6. https://doi.org/10.17239/L1ESLL-2018.18.04.01

Bonset, H., & Hoogeveen, M. (2010). *Taalbeschouwing. Een inventarisatie van empirisch onderzoek in basis- en voortgezet onderwijs* [Language contemplation. An inventory of empirical research in primary and secondary education]. Enschede: SLO.

Camps, A., & Fontich, X. (2019). Teachers' concepts on the teaching of grammar in relation to the teaching of writing in Spain: A case study. *L1-Educational Studies in Language and Literature, 19*, 1-36. https://doi.org/10.17239/L1ESLL-2019.19.02.02

Chomsky, N. (1965). *Aspects of the theory of syntax.* MIT Press. https://doi.org/10.21236/AD0616323

Coppen, P.-A., Van Rijt, J., Wijnands, A., & Dielemans, R. (2019). Vakdidactisch onderzoek naar grammaticaonderwijs in Nederland [Educational research on grammar education in the Netherlands]. *Tijdschrift voor Nederlandse Taal- en Letterkunde, 135*(2), 85-99. https://hdl.handle.net/2066/206

Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal, 18,* 391-404. https://doi.org/10.1207/s15326934crj1803_13

Curriculum.nu (2019). *Conceptvoorstellen leergebied Nederlands* [Concept proposals subject area Dutch], 2019.

Delarue, S. (2016). Bridging the policy-practice gap: how Flemish teachers' standard language perceptions navigate between monovarietal policy and multivarietal practice. Doctoral Dissertation. Ghent University, Ghent.

Denham, K. (2020). Positioning students as linguistic and social experts: teaching grammar and linguistics in the United States. *L1-Educational Studies in Language and Literature 20*, 1-16. https://doi.org/10.17239/L1ESLL-2020.20.03.02

Denham, K., & Lobeck, A. (2010). (Eds.), *Linguistics at school: language awareness in primary and secondary education*. Cambridge University Press. https://doi.org/10.1017/CBO9780511770791

Depaepe, F., De Corte, E., & Verschaffel, L. (2016). Mathematical epistemological beliefs. In J.A. Green, W.A. Sandoval, & I. Bråten. (Eds.), *Handbook of epistemic cognition* (pp. 147-164). Routledge.

Derewianka, B. (2012). Knowledge about language in the Australian curriculum: English. *Australian Journal of Language and Literacy*, *35*(1), 127-146.

Dielemans, R., & Coppen, P.-A. (2021). Defining linguistic reasoning. Transposing and grounding a model for historical reasoning to the linguistic domain. *Dutch Journal of Applied Linguistics*, *9*(1), 182-206. https://doi.org/10.1075/dujal.19038.die

Elby, A., Macrander, C., & Hammer, D. (2016). Epistemic cognition in science. In Green, J.A., Sandoval, W.A., & Bråten, I (Eds.), *Handbook of epistemic cognition* (pp. 113-127). Routledge.

Elsner, D. (2020). Grammar is irrelevant. The role of epistemological beliefs in students' learning success. In A. Trotzke, & T. Kupisch (Eds.), *Formal linguistics and language education. New empirical perspectives* (pp. 69 – 90). Cham: Springer. https://doi.org/10.1007/978-3-030-39257-4_5

Exley, B., & Mills, K. A. (2012). Parsing the Australian English curriculum: Grammar, multimodality and cross-cultural texts. *The Australian Journal of Language and Literacy, 35*(2), 192-205.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics (4th ed.).* Sage.

Fontich, X., & García-Folgado, M. (2018). Grammar instruction in the Hispanic area: The case of Spain with attention to empirical studies on metalinguistic activity. *L1-Educational Studies in Language and Literature, 18*, 1-39. https://doi.org/10.17239/L1ESLL-2018.18.04.02

Hofer, B.K. (2002). Personal epistemology as a psychological and educational construct: An introduction. In B.K. Hofer, & P.R. Pintrich (Eds.), *The psychology of beliefs about knowledge and knowing* (pp. 3-14). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410604316

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about Knowledge and knowing and their relation to learning. *Review of Educational Research, 67,* 88–140. http://dx.doi.org/10.3102/00346543067001088

Honda, M., & O'Neil, W., (2008). *Thinking linguistically: A scientific approach to language*. Wiley-Blackwell.

Huang, F. L. (2017). Conducting multilevel confirmatory factor analysis using R. Retrieved from http://faculty.missouri.edu/huangf/data/mcfa/MCFA%20in%20R%20HUANG.pdf

Hudson, R. (1999). Grammar-teaching is dead – NOT! In R. Wheeler (Ed.), *Language alive in the classroom* (pp. 101–112). Greenwood.

Hudson, R. (2004). Why education needs linguistics (and vice versa). *Journal of Linguistics, 40*, 105-130. https://doi.org/10.1017/S0022226703002342

Hulshof, H. (2014). De 'triviale' herkomst van het redekundig ontleden. Uit de geschiedenis van de school grammatica in Nederland [The 'trivial' origins of sentence parsing. From the history of school grammar in the Netherlands]. *Levende Talen Magazine, 101*(7), 22-25.

Jaarsveld, S., Lachmann, T., & Van Leeuwen, C. (2012). Creative reasoning across developmental levels: Convergence and divergence in problem creation, *Intelligence, 40*(2), 172-188. https://doi.org/10.1016/j.intell.2012.01.002

Kienhaus, D, Ferguson, L., & Stahl, E. (2016). Diverging information and epistemic change. In J.A. Green, W.A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 318-330). Routledge.

King, P.M. (2009). Principles of development and developmental change underlying theories of cognitive and moral development. *Journal of College Student Development, 50*(6), 597-620. https://doi.org/10.1353/csd.0.0104

King, P.M., & Kitchener, K.S. (1994). *Developing reflective judgement: Understanding and promoting intellectual growth and critical thinking in adolescents and adults.* Jossey-Bass Publishers.

King, P.M., & Kitchener, K.S. (2002). The reflective judgment model: Twenty years of research on epistemeic cognition. In B.K. Hofer, & P.R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 37-61). Lawrence Erlbaum.

King, P.M., & Kitchener, K.S. (2004). Reflective Judgement: Theory and Research on the Development of Epistemic Assumptions Through Adulthood. *Educational Psychologist, 39*(1), 5-18. https://doi.org/10.1207/s15326985ep3901_2

Knop, L., & Van Laere, A. (2017). Taal actief! [Language active!]. In S. Vanhooren, & A. Mottart (Eds.), *33-ste conferentie onderwijs Nederlands* [33st Conference Education Dutch] (pp. 246-248). Scribis.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15*(3), 309–328. https://doi.org/10.1016/S0885-2014(00)00030-7

Lybaert, C. (2015). De ideologische positie van Algemeen Nederlands in Vlaanderen. De standaardiseringsideologie onder druk? [The ideological position of Standard Dutch in Flanders. Standardization ideology under pressure?] *Taal en Tongval, 67*(1), 97-134. https://doi.org/10.5117/TET2015.1.LYBA

Maggioni, L. (2010). *Studying epistemic cognition in the history classroom: Cases of teaching and learning to think historically.* Doctoral dissertation. University of Maryland, Maryland.

Maggioni, L., Alexander, P., & VanSledright B. (2004). At a crossroad? The development of epistemologi-cal beliefs and historical thinking. *European Journal of School Psychology, 2*(1-2), 169-197

Maggioni, L., VanSledright, B., & Alexander, P. (2009). Walking on the borders: A measure of epistemic c-ognition in history. *The Journal of Experimental Education, 77*(3), 187-214. https://doi.org/10.3200/JEXE.77.3.187-214

Mason, L. (2016). Psychological perspective on measuring epistemic cognition. In J.A. Green, W.A. Sando-val, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 375-392). Routledge.

Meesterschapsteam-Nederlands (2018). *Visie op de toekomst van het schoolvak Nederlands* [Vision on t-he future of the school subject Dutch]. *Scribbr.* https://nederlands.vakdidactiekgw.nl/.

Mierwald, M., Lehman, T., Seiffert, J., & Brauch, N. (2017). Fragebogen auf dem Prüfstand - Ein Beitrag z-ur Erforschung und Weiterentwicklung eines bestehenden Instruments zur Erfassung episte-meologischer Überzeugungen in der Domäne Geschichte [Questionnaire on the test bench - A contribution to the research and further development of an existing instrument for the asse-ssment of epistemological beliefs in the domain of history]. In M. Waldis, & B. Ziegler (Eds.), *Forschungswerkstatt Geschichtsdidaktik 15. Beiträge zur Tagung 'Geschichtsdidaktik empirisch 15' [Research workshop history didactics 15. Contributions to the conference 'Geschichtsdidak-tik empirisch 15']* (pp. 177-190). Hep, der Bildungsverlag.

Muis, K.R. (2004). Personal epistemology and mathematics: A critical review and synthesis of research. *R-eview of Educational Research, 74*(3)*,* 317-377. https://doi.org/10.3102/00346543074003317

Muis, K.R. (2007). The role of epistemic beliefs in self-regulated learning. *Educational Psychologist, 42*(3)*,* 173-190. https://doi.org/10.1080/00461520701416306

Muis, K.R., Bendixen L.D., & Haerle, F. C. (2006). Domain-generality and domain-specificity in personal e-pistemology research: Philosophical and empirical reflections in the development of a theo-retical framework. *Educational Psychology Review, 18*(3)*,* 3-54. https://doi.org/10.1007/s10648-006-9003-6

Roberts, L. D. (2007). Equivalence of electronic and off-line measures. In R. A. Reynolds, R. Woods, & J. D. Baker (Eds.), *Handbook of research on electronic surveys and measurements* (pp. 97–103). Idea Group. https://doi.org/10.4018/978-1-59140-792-8.ch010

Stoel, G.L. (2017). *Teaching towards historical expertise. Developing students' ability to reason causality in history.* Doctoral dissertation. University of Amsterdam, Amsterdam.

Stoel, G., Logtenberg, A., Wansink, B., Huijgen, T., Van Boxtel, C., & Van Drie, J. (2017) Measuring episte-mological beliefs in history education: An exploration of naïve and nuanced beliefs. *Interna-tional Journal of Educational Research, 83*, 120–134. https://doi.org/10.1016/j.ijer.2017.03.003

Tabachnick, B. & Fidell, L. (2013). *Using multivariate statistics (Sixth Ed.).* Pearson Education.

Trousdale, G (2010), Knowledge about language in the English classroom: a Scottish perspective. In T. Lo-cke (Ed.), *Beyond the grammar wars: A resource for teachers and students on developing lan-guage knowledge in the English/literacy classroom (*pp. 73-87). Routledge.

Van den Broek, E.W.R. (2020). *Language awareness in foreign language education. Exploring teachers' b-eliefs, practices and change processes*. Doctoral dissertation. Radboud University Nijmegen, Nijmegen. https://doi.org/10.1080/09571736.2019.1688857

Van der Aalsvoort, M. (2016). *Vensters op vakontwikkeling. De betwiste invoering van taalkunde in het examenprogramma Nederlands havo/vwo (1988-2008)* [Windows on subject development. Th-e contested introduction of linguistics in the exam programme Dutch havo/vwo (1988–2008)]. Doctoral dissertation. Radboud Universiteit, Nijmegen.

Van Gelderen, A. (2010). Does explicit teaching of grammar help students to become better writers? Ins-ights from empirical research. In T. Locke (Ed.). *Beyond the grammar wars. A resource for tea-chers and students on developing language knowledge in the English/literacy classroom* (pp. 109–128). Routledge.

Van Rijt, J.H.M. (2020). *Understanding grammar. The impact of linguistic metaconcepts on L1 grammar education*. Doctoral dissertation. Radboud University Nijmegen, Nijmegen. https://doi.org/10.1080/02671522.2020.1784258

Van Rijt, J., & Coppen, P.-A. (2017). Bridging the gap between linguistic theory and L1 grammar education - experts' views on essential linguistic concepts. *Language Awareness, 26*(4), 360-380. https://doi.org/10.1080/09658416.2017.1410552

Van Rijt, J., Wijnands, A., & Coppen, P.-A. (2019). Dutch teachers' beliefs on linguistic concepts and reflective judgement in grammar teaching. *L1-Educational Studies in Language and Literature, 19, 1-28.* https://doi.org/10.17239/L1ESLL-2019.19.02.03

VanSledright, B., & Maggioni, L. (2016). Epistemic cognition in history. In J.A. Green, W.A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 128-146). New York: Routledge.

VVKSO. (2014). *Leerplan secundair onderwijs, Nederlands, derde graad aso-kso-tso* [Curriculum Secondary education, Dutch, third degree general secondary education – artistic secondary education – technical secondary education]. Brussel: VVKSO (D/2012/7841/0001).

Wijnands, A., Van Rijt, J., & Coppen, P.-A. (2021). Learning to think about language step by step. A pedagogical template for the development of cognitive and reflective thinking skills in L1 grammar education. *Language Awareness, 30*(4)*,* 317-335. https://doi.org/10.1080/09658416.2021.1871911

Wijnands, A., Van Rijt, J.H.M., Stoel, G.L., & Coppen, P.-A. (forthcoming). Balancing between uncertainty and control. Teaching reflective thinking about language in the classroom.

Yuen Lie Lim, L-A (2011). A comparison of students' reflective thinking across different years in a problem-based learning environment. *Instructional Science, 39*(2), 171-188. https://doi.org/10.1007/s11251-009-9123-8

APPENDICES

Appendix A. Protocol

(Note: Only text fragments of the protocol concerning the questionnaire has been given here. For instance, text fragments about declaration of consent have been removed)

Thank you very much for participating with your students in our research. In order to generate the data as validly as possible, I would like to ask you the follow the following protocol.

- The teacher will handout the questionnaire.
- Students will get about 10 minutes for filling in the questionnaire. Afterwards, the teacher will collect the questionnaires.
- It is not allowed to communicate about the questionnaire when filling in the questionnaire.
- The teacher is not allowed to answer any questions of the students about the questionnaire.
- After collecting in the questionnaires, the teacher is free to discuss what the students had to accomplish.

This protocol is strict in order to create comparable test conditions in different classrooms with different teachers. If you have any questions about this protocol, please contact us. If you noticed something important during the test, please report this as well.

Appendix B. Conversion of history phrases (Stoel et al., 2017) into linguistic phrases

History phrase used by Stoel et al. (2017)	Converted linguistic phrase
History / A history account	The analysis of words and sentences
History inquiry	Analyzing words and sentences
Different stories about the past	Analyzing the form and meaning of a word or sentence in different ways
Past	Form and meaning
To know what happened in the past	To know the correct form and meaning of a word or sentence
Historical events	Language intuitions
The same claim about a historical event	the same judgment about the analysis of a word or sentence
To give a same explanation for an event	To analyze words and sentences in a similar way
Interpretation	Language intuition
Causes	Language rules and language intuitions
Complete evidence	To follow all steps
Eyewitnesses	People's judgment of the grammaticality of a word or sentence.
Multiple perspectives on the past	Different perspectives on form and meaning into account (What do you think? What do other people think? What are the rules?)

Appendix C. Experts' scores on the questionnaire

		Experts						
		1	2	3	4	5	6	7
	Prereflective							
1	The goal of analyzing words and sentences is to find the correct solution	2	2	1	2	5	1	4
3	When two persons have the same judgment about the analysis of a word or sentence, you know it is true.	3	2	1	2	2	2	2
8	Linguistic experts will analyze words and sentences in the same way, when they consult the same linguistic resources.	3	2	1	2	2	5	4
9	The basis of one's language intuitions are the prescriptive language rules.	2	2	1	1	1	4	2
16	When some people consider a word or sentence as ungrammatical, it is impossible to analyze this word or sentence in a proper way.	2	2	2	1	1	2	1
17	It is not possible to analyze words and sentences adequately when these words and sentences can be viewed differently.	2	2	2	1	2	2	2
18	Linguistic experts will probably have the same answers on questions about form and meaning.	4	2	2	1	5	5	2
20	A word or sentence can only be analyzed correctly when you have gone through all the steps.	4	2	4	2	3	4	2
21	When a language rule is written in your textbook, you can be nearly certain that it is true.	4	2	4	1	2	2	4
23	When the form and meaning of a word or sentence can be analyzed in different ways, only one can be correct.	2	2	1	1	1	1	2
	Quasi-reflective							
7	Analyses of words or sentences are mainly opinions.	2	2	1	2	1	2	2
10	Since your language intuitions cannot be observed in a reliable way, you will never know whether those intuitions are correct.	2	4	4	4	1	5	3
12	Analyses of words or sentences are largely opinions of linguistic experts.	2	2	3	1	2	2	1
19	You can never know for certain the correct form and meaning of a word or sentence.	3	5	4	3	2	2	5
25	In a thorough analysis, all the causes of your own language intuitions will become clear.	4	4	3	4	3	3	3

Reflective

2	There are various methods to prove the correctness of an analysis of words and sentences.	5	5	4	5	5	5	6
4	When analyzing words and sentences you must learn to deal with conflicting information.	5	5	6	5	5	5	6
5	The analysis of words and sentences is a good way to develop your inquiry skills	5	6	4	6	5	5	5
6	When analyzing words and sentences it is important that you learn to support your reasoning with evidence.	5	6	4	6	5	5	6
11	When consulting sources it is important to check the origin of the source.	5	6	6	6	5	6	5
13	New explanations for language intuitions will always be invented.	4	3	4	5	4	5	3
14	In the analysis of words and sentences your own language intuitions are important.	5	4	5	6	5	5	2
15	A good analysis of words and sentences must take different perspectives on form and meaning into account (What do you think? What do other people think? What are the rules?).	5	6	4	6	3	5	5
22	The analysis of words and sentences is a critical search for form and meaning.	5	6	6	6	5	5	4
24	Even when consulting the same sources, linguistic experts often arrive at different analyses.	4	5	6	5	3	5	4
26	In many cases, linguistic experts will continue to discuss language rules and language intuitions.	4	5	5	5	2	5	5