‘X is the odd one out, because the other two are more about the farmland’ – Dutch L1 student teachers’ struggles to reason about grammar in odd one out tasks

Jimmy H.M. van Rijt a,b,c, Hans Hulshof c, Peter-Arno J.M. Coppen d

a Fontys University of Applied Sciences, Department of Teacher Education, P.O. box: 558, 6130 AN Sittard, the Netherlands
b Radboud University Nijmegen, Department of Linguistics, Centre for Language Studies, P.O. box 9103, 6500 HD Nijmegen, the Netherlands
c Leiden University, Faculty of Humanities, Centre for the Arts in Society, ICLO, P.N. van Eyckhof 4, 2311 BV Leiden, the Netherlands
d Radboud University, Radboud Teachers’ Academy, P.O. box 9103, 6500 HD Nijmegen, the Netherlands

A R T I C L E  I N F O

Keywords:
L1 grammar
Grammatical reasoning
Knowledge about language
Teacher training
Odd one out tasks

A B S T R A C T

Grammatical knowledge is an important part of L1 language education. Nevertheless, teachers find it challenging to convey an in-depth understanding of grammar to their students. Previous research suggests that understanding might be stimulated by focusing on grammatical reasoning. The current mixed-methods study explores the grammatical reasoning of 108 Dutch L1 student teachers’ in odd one out tasks, showing that student teachers struggle with such reasoning tasks. A multilevel regression analysis indicates that their level of grammatical understanding as measured by a Test of Grammatical Understanding (TGU) and the elaborateness of student argumentation significantly predict the quality of their grammatical reasoning. Student teachers' performances were also compared to 14 year old pre-university students’ performances (N = 120). Contrary to what was hypothesized, senior student teachers did not manage to outperform junior student teachers, nor did student teachers outperform pre-university students. The paper discusses plausible reasons for these findings and explores how teacher education might need to shift focus to better develop student teachers’ grammatical reasoning skills.

1. Introduction

While debates about the role of explicit L1 grammar teaching have been numerous over the past decades and continue to persevere until the present day, there appears to be a growing consensus that knowledge about language is an important educational goal (Fontich & Camps, 2014; Locke, 2010; Rattya, Awramiuk, & Fontich, 2019). Knowledge about Language (Kal.) has re)claimed an important position in many educational jurisdictions, including Anglophone countries (Locke, 2010; Macken-Horakí, Love, & Horakí, 2018; Myhill, 2018), Francophone countries (Boivin, 2018), Spanish speaking regions (Fontich & García-Folgado, 2018) and Central-European regions (e.g., Awramiuk & Smytna, 2019; Stepánik, 2019). In the Netherlands, where the current study is set, a large curriculum reform is underway in which Kal. seems likely to be attributed a more prominent place in the curriculum (Curriculum.nu, 2019). An important part of Kal. concerns grammar or syntax: knowing how sentences and phrases are structured, what they mean and how form and meaning relate to one another. Much research has been dedicated to investigate how grammatical knowledge might impact on the development of literacy, particularly writing (Andrews, 2005; Andrews, 2010; Fontich & Camps, 2014; Gordon, 2005; Graham & Perin, 2007; Locke, 2010; Myhill, 2018). And while influential studies such as those by Myhill, Jones, Lines, and Watson (2012), Jones, Myhill, and Bailey (2013) and Myhill (2018) have demonstrated that explicit knowledge about grammar can positively impact on students’ writing development when taught in context, it can also be considered a valuable goal in its own right, as knowledge about one of the most critically important parts of human culture and society (Hudson, 2004; Van Rijt, 2020). Whatever the reason for teaching grammar (literacy development, understanding the language system or a combination of these), the question remains how explicit grammatical knowledge can best be taught to foster students’ in-depth grammatical understanding, as this is one of the greatest challenges for language teachers (Andrews, 1997; Myhill, 2000; Myhill, 2003; Sangster, Anderson, & O’Hara, 2013). How do they...
ensure their students can really comprehend grammatical constructions or phenomena beyond a shallow level, and if a literacy-related perspective towards grammar teaching is maintained, how can such knowledge be transferred to reading and writing (Fontich, 2016; Watson & Newman, 2017)? Recent research has demonstrated that grammatical understanding can be improved by short interventions targeting underlying linguistic metaconcepts, both at the university level (Van Rijt, De Swart, Wijnands, & Coppen, 2019) and at the secondary school level (Van Rijt, Wijnands, & Coppen, 2020a). In such interventions, the aim is to first establish an understanding of a larger part of the language system (a metaconcept) before refining that understanding with the grammatical concepts that are subordinate to the metaconcept. For example, the metaconcept of valency conveys the insight that verbs require and select roles (‘arguments’), which are needed for the verb to be properly understood. These roles correspond with syntactic functions, such as subject, direct object and indirect object. Some verbs require only one role (e.g., to walk, to sleep), whereas others require two (to read, to build) or three (to give, to donate). Thus, depending on the valency of the verb, some sentences contain objects (e.g., when the main verb is to read, to give), whereas others do not (e.g., when the main verb is to walk). Metaconcepts such as valency can therefore facilitate the understanding of grammatical concepts such as subject and objects, and they have been shown to substantially enhance both university and secondary school students’ ability to reason about unknown grammatical problems, which might be indicative of their increased grammatical understanding. These studies had effect sizes ranging between .46 (Van Rijt et al., 2020a) and .62 (Van Rijt, De Swart et al., 2019), which is substantial for short educational interventions (Calin-Jageman & Cumming, 2018). A subsequent quasi-experimental study with switching replications (Van Rijt, 2020, ch. 6) found similar effects on pre-university students’ level of grammatical understanding as measured by a Test of Grammatical Understanding (TGU, see section 2.2.2). Apart from introducing students to linguistic metaconcepts, the interventions encouraged reasoning about grammatical concepts and metaconcepts (cf. Dielemans & Coppen, 2021; Honda & O’Neil, 2007). In part, this was stimulated by odd one out tasks, which have been shown to be effective in enhancing historical reasoning in history education (Havekes, 2015). (See Table 2 for grammar examples.) In an odd one out task, students are presented with several grammatical units (e.g., verbs) and are invited to argue which of these is the odd one out (i.e., different from the other ones), and, most importantly, why. In a properly designed odd one out task, each of the options could potentially be ruled out, to ensure that the answer is not clear-cut and critical thinking is stimulated. The main aim of such tasks is thus to stimulate grammatical reasoning that demonstrates an understanding of the subject matter. Since such tasks have been used successfully to improve secondary school students’ historical reasoning (Havekes, 2015) as well as their grammatical reasoning (Van Rijt, 2020), it seems that such tasks are well suited for stimulating grammatical understanding at the secondary school level and beyond. A recent national survey among Dutch language teachers in which grammar teaching practices were investigated has shown that most Dutch language teachers favour types of grammar teaching in which in-depth understanding and grammatical reasoning are encouraged (Van Rijt, Wijnands, & Coppen, 2020b), although the same study also demonstrated that most teachers’ actual practices are much more traditional, i.e., not involving underlying metaconcepts and not being focused on grammatical reasoning, but rather on parsing decontextualized sentences (cf. Van Gelderen, 2010). The question thus seems to be how teachers can be moved towards more metacausal grammar lessons. It may be that the reason they do not engage in them is (partly) due to their own grammatical insecurities. Previous studies have shown that most teachers are often anxious when they have to teach grammar (Cajkler & Hislam, 2002; Giovanelli, 2015), and that their own Kal is typically underdeveloped (Alderson & Hudson, 2013; Sangster et al., 2013; Van Rijt, Wijnands, & Coppen, 2019; Watson, 2012). At the same time, teachers’ own level of grammatical understanding can predict how effectively they can teach grammar, especially in the context of writing (Myhill, Jones, & Watson, 2013). In other words: the more grammatical knowledge they possess, the more their students will benefit from grammatical interventions. Examining how well teachers can deal with grammatical reasoning tasks thus seems to be a prerequisite for improving grammar education accordingly. At the same time, it should be acknowledged that subject knowledge alone is a necessary but insufficient condition, as pedagogic subject knowledge (i.e., ‘how to teach effectively’) is arguably an even more important aspect of teacher knowledge (Myhill et al., 2013). It might be argued that a good place to start investigating how well teachers can reason about grammatical problems is in teacher education programs. After all, student teachers’ beliefs towards new forms of grammar teaching can be shaped more easily if they have not yet been fully immersed into classroom practices (Grass, 2018), since research shows that teacher beliefs tend to be influenced heavily by traditional practices. This paper therefore investigates how well Dutch language student teachers can reason about L1 grammar, specifically in odd one out tasks. Given the importance of the teacher in teaching grammar (Myhill et al., 2013), future teachers should be able to outperform secondary school students in this ability. We will therefore also examine how student teachers reason compared to 14 year-old pre-university students. Before explaining our research methods, we will first briefly provide some necessary context about the Dutch teacher education system.

1.1. Teacher education for Dutch language teachers

In the Netherlands, there are two programs for Dutch language teacher training, for training students to teach at the secondary school level. One is the university program, in which a regular study of Dutch language and literature is completed by a specific (sometimes integrated) teacher training. This five year program leads to a first degree teacher certificate, in combination with a Master of Education degree (MEd). Such teachers are licensed to teach at all levels of secondary education. The other route is a university of applied sciences program (in Dutch ‘hbo’), which is an integrated teacher training and Dutch language and literature study, leading to a Bachelor of Education degree (BEd) after four years, in combination with a second degree teacher certificate. BEd licensed teachers are only allowed to teach in the lower classes of secondary education. Older students who are making a career switch into education can take the same program, which usually takes two years instead of four (parttime)². The BEd program can be extended with a 2–3 year master program, leading to a first degree teacher certificate, and an MEd degree. The university program is usually followed by students from pre-university secondary education (vwo), whereas the hbo program is the usual route for students from higher vocational secondary education (havo). In the hbo program, where this study takes place, BEd students are trained in both syntax and general linguistics within the first two years of their education. Their training encompasses knowledge and analytical skills related to traditional grammar (classical parts of speech (e.g., adjective, noun, verb) and phrases (e.g., subject, direct object, adverbial), followed by more modern grammatical (meta) concepts (e.g., modality, predication, recursion, valency – see also Van Rijt & Coppen, 2017). In the final two years of their program, they also receive courses on how to teach grammar or language awareness. Upon

² The reason that the four year fulltime program can also be taken in two years parttime is that those students in most cases have completed another program that overlaps for a large part with the teacher education program, in particular the pedagogical part, thus enabling them as lateral entrants to obtain exemptions for general parts of the teacher training program based on previously acquired competencies (RPAC).
completing their teacher education program, student teachers should thus possess sufficient knowledge to be able to teach grammar effectively. Throughout their time, starting in year 1, they will have continuous internships to prepare them for their teaching jobs. MEd students’ knowledge about language is expanded upon by (more) advanced master’s courses in linguistics. No special attention is given to grammar teaching at MEd programs, since grammar is usually only taught in the lower levels of secondary education in the Netherlands, until students are about 14 years old ( Van der Aalsvoort, 2016 ), and rarely in the upper classes ( Meestringa & Ravesloot, 2013 ), although schools are allowed to do so if they choose. Hence, students from the upper classes of secondary education are rarely taught explicit grammar. It is important to note that the school subject of Dutch language and literature is in itself quite similar for higher vocational and pre-university students, albeit that more is expected from pre-university students, especially with regard to their level of knowledge and critical thinking ability. In addition, pre-university education takes a year longer than higher vocational education (six vs. five years, respectively). Pre-university students are thus likely to possess more grammatical knowledge than their higher vocational counterparts, especially considering that some of the pre-university students have been taught Greek and Latin in addition to Dutch and some modern foreign languages, in which grammar is strongly emphasized. The official curriculum, however, does not demand that pre-university students are taught more grammar (cf. Mei jerink et al., 2008 ), although as a rule of thumb, the higher the pupils’ level, the more likely it is they have come across more grammar in educational practice. Although grammar is usually only taught in the lower classes of secondary education, it should be expected that L1 student teachers have a more profound understanding of grammar than students from secondary education.

1.2. Research questions

To reiterate, in the current study we investigated how well Dutch language student teachers from universities of applied sciences can reason about grammar. More specifically, we aimed to investigate the following issues:

1. What are the characteristics of student teachers’ grammatical reasoning in odd one out tasks?
2. How does the grammatical reasoning of student teachers compare to the grammatical reasoning of 14 year old pre-university students?
3. To what extent are there differences in reasoning quality between junior and senior student teachers?
4. To what extent does student teachers’ general level of grammatical understanding predict the quality of their reasoning about grammar?

It is important to note that in this study, ‘grammatical reasoning’ and ‘grammatical argumentation’ are used as synonyms throughout. We hypothesize that student teachers will have trouble reasoning about grammar, as they typically have not come across reasoning tasks in their own grammar education a lot. However, it is also reasonable to hypothesize that student teachers will generally outperform 14 year old pre-university students, as it can be assumed that student teachers are more motivated and better trained to handle grammar. To the best of our knowledge this study is the first to empirically explore how well student teachers can deal with grammatical odd one out reasoning tasks.

2. Method

This study follows a mixed-method design, examining both quantitative and qualitative data.

2.1. Participants

In this study, 108 student teachers from 8 different Universities of Applied Sciences in the Netherlands participated. Each of these universities supplied between 8.3 and 32.4 % of the respondents. Teacher education institutes (hbo only) were contacted by the first author to invite their student teachers to fill in a questionnaire online via Qualtrics. All of the institutions for teaching education (9 in total) were contacted; 8 of them replied positively. There are no substantial quality differences between these institutions. All of them have been positively accredited by The Accreditation Organisation of the Netherlands and Flanders (NVAO). In the first section of the questionnaire, student teachers were asked to report general demographic data, such as age, gender and the institution to which they were associated. Active permission to use student teachers’ data anonymously for scientific research was obtained. Table 1 lists details about participants’ characteristics.

Of these student teachers, only the MEd students had actual working experience as Dutch teachers at the secondary school level at the time of the investigation. The rest of the student teachers should be regarded as pre-service teachers, whose highly limited teaching experience has come from supervised internships only. For the purpose of comparing students teachers’ reasoning with the reasoning of secondary school students, we used data from a previous study (Van Rijt, 2020, ch. 6). As part of a short metaconceptual intervention (4 lessons) in which they were taught about valency, a group of pre-university students completed the same odd one out tasks during the intervention as the student teachers. 120 pre-university students from 5 different secondary schools (M age = 14.04, SD = 0.45) completed the tasks. Of these pre-university students, 62 were male and 58 were female. For extensive details on the other activities of the 14 year olds in this intervention, see Van Rijt (2020, ch. 6).

2.2. Questionnaire

2.2.1. Odd one out tasks

After having provided personal information, student teachers were presented with an explanation of the grammatical odd one out tasks. They were provided with an example unrelated to grammar to illustrate what was expected of them (given these three animals cat, crocodile, lion - which one is the odd one out?). Student teachers were encouraged to provide arguments for each option they felt could be the odd one out, and they were asked to come up with the best grammatical arguments possible. This means they could provide arguments for 1, 2 or 3 alternatives, depending on how many options they felt could be excluded. Student teachers were prompted to use the following sentence to help guide their argumentation: ‘X is the odd one out, because the other two …’, forcing them not only to address the oddness of the odd one out, but also to search for a similarity between the other options. It is known from research into historical reasoning that such a formulation can help deepen the arguments ( Havekes, 2015 ). The student teachers were then given the odd one out tasks in Table 2 (in Dutch), for which they had not been trained specifically. Both the order of the tasks and the alternatives from which the student teachers could choose were randomized to rule out any order effects. No word limits were imposed so as not to restrict

| Track                  | Total N | Mean age (SD) | N male | N female |
|------------------------|---------|---------------|--------|----------|
| BEd fulltime           | 61      | 21.05 (2.24)  | 13     | 48       |
| BEd parttime           | 36      | 39.58 (11.52) | 7      | 29       |
| MEd parttime           | 11      | 36.45 (8.26)  | 1      | 10       |
| Total                  | 108     | 28.8 (11.49)  | 21     | 87       |

Table 1 Student teacher characteristics (N = 108) per teacher trainer track.
After these tasks, student teachers were asked to reflect on some statements on a five point Likert scale, ranging from 'Strongly disagree' to 'Strongly agree'. Topics included the amount of effort put into the task, confidence about their performance, their willingness to use these tasks in their own lessons, and their previous familiarity with the type of the task.

2.2.2. Test of Grammatical Understanding (TGU)

To measure students teachers’ level of grammatical understanding, they were asked to complete the Test of Grammatical Understanding (TGU) – see Appendix A. This test was developed for measuring grammatical insight for secondary school students. The test consisted of twelve multiple choice questions in which a grammatical problem was described. The grammatical problems revolved around the metaconcept of valency (also covering the traditional phrases and parts of speech related to valency), although the term valency was never explicitly used. Instead, the questions were designed in such a way that they could be answered by anyone who had an understanding of valency, even if they did not know the term. Each question was provided with four multiple choice alternatives, and student teachers were tasked with choosing the best alternative. This construct was closely informed by previous research (Van Rijt et al., 2020a), and by the literature about understanding (De Regt, 2009), which states that understanding comes in degrees (Baumberger, Beisbart, & Brun, 2016; Baumberger, 2019). The multiple choice items were thus meant to reflect different degrees of understanding. One of the alternatives conveyed the complete insight, with appropriate grammatical terminology, for which participants could receive 2 points if chosen. Another alternative conveyed partial grammatical insight, so participants could receive 1 point for choosing this alternative. The third alternative used a grammatical concept ‘blindly’, i.e., in a wrong way, or as irrelevant for the case at hand. Student teachers who chose this alternative received 0 points. The fourth alternative, finally, was an intuitive answer to the grammatical problem, in which no grammatical concepts were used at all. This answer also yielded 0 points. Table 3 illustrates an example of a question (with alternatives) from the TGU.

The TGU was developed by two linguists and carefully pretested in secondary education. Additional tests among teachers with linguistic expertise showed that the test is able to distinguish between more or less grammatically informed individuals. See Van Rijt (2020) for more details about the TGU. In Van Rijt (2020, ch. 6), three versions of the TGU were administered. In the current study, one of these versions was randomly chosen for all student teachers. The order of all questions from the TGU and the answers were randomized. Since the TGU was used in previous research to measure pre-university students’ progress in grammatical understanding after a metaconceptual intervention, we could compare pre- and post-intervention scores for the pre-university students, and compare these results to the performance of our student teachers. This way, we were able to gain a deeper understanding of the student teachers’ level of grammatical understanding.

2.3. Data analysis

2.3.1. Qualitative analysis

The pre-university students’ and student teachers’ reasonings were analyzed qualitatively, using a coding scheme from previous research (Van Rijt, 2020) that was developed by two coders (one of whom was the first author of this paper). The coding scheme was based on the constant comparison method (Corbin & Strauss, 2015; Wellington, 2000) in which the coders first coded the data individually, and then developed a joint coding scheme that was continuously refined by going back to the data several times and by discussing coding issues until a full agreement on the relevant coding was reached. The current study adopts this coding scheme to allow for comparisons across student populations. These codes provide insights into the characteristics of students’ grammatical reasoning.

2.3.2. Quantitative analysis

2.3.2.1. Scoring students’ grammatical reasoning. To gain an idea of how well students teachers’ grammatical reasoning was, two experienced teacher educators with expertise in linguistics independently rated each reasoning holistically on a 10 point Likert scale. A 10 point scale was chosen because such a scale is common in the Dutch educational system, making the task more natural for the raters. To determine interrater reliability, we calculated a two-way mixed, absolute, average-measures intra-class correlation (ICC, cf. McGraw & Wong, 1996), which was in the excellent range for both Task 1 (ICC = .94) and Task 2 (ICC = .95) – see Cicchetti (1994). These scores were then normalized by calculating Z scores for each rater, to account for scale variance between the two raters. These Z scores served as the input for the multilevel analyses.
Studies in Educational Evaluation 70 (2021) 101007

5

model as an additional predictor (see 3.1). We thus used a step-up analysis in which the total number of words was incorporated into the test.

quality score of the raters, we also ran a third multilevel regression significant correlation between the number of words per task and the dent samples T-tests. Because an initial exploration of the data showed a fourth year fulltime BEd students, as well as second year parttime BEd reasoning ability, we divided these student teachers into three cate differences between junior and senior student teachers in grammatical

described in section 2.3.2.2.

2.3.2.2. Statistical analyses. In order to investigate whether there were differences between junior and senior student teachers in grammatical reasoning ability, we divided these student teachers into three categories, based on similarities in their teacher training program: (1) level 1 student teachers, which were formed by students in their first two years of a fulltime BEd track and the first-year students of a part-time BEd track; (2) level 2 student teachers, which were formed by third and fourth year fulltime BEd students, as well as second year parttime BEd students; (3) level 3 student teachers, which were formed by the masters’ students. We refer to this variable as student teacher level. To investigate the effect of student teacher level, we first ran a multilevel regression model controlling for the effect of educational institute in which we explored the effect of student teacher level on reasoning scores (Z-scores) for the odd one out tasks. Next, we included TGU scores to the model to see whether student teachers’ general understanding of grammar could predict their reasoning scores better than student level alone. In this process, we took student teacher level and the TGU score into account as fixed effects (i.e., we assumed that the effects of student teachers’ level and TGU scores were similar for all student teachers) and institution as a random effect (i.e., we assumed that odd one out scores varied across institutions). For a better interpretation of the TGU scores of the student teachers, we also examined differences in TGU scores between student teachers and secondary school students via independent samples T-tests. Because an initial exploration of the data showed a significant correlation between the number of words per task and the quality score of the raters, we also ran a third multilevel regression analysis in which the total number of words was incorporated into the model as an additional predictor (see 3.1). We thus used a step-up building strategy (West, Welch, & Galecki, 2007), and determined which of these three models was best by comparing the -2 log likelihood values of the extended model compared to the previous model, using a $\chi^2$ test.

3. Results

3.1. Characteristics of grammatical reasoning (qualitative analysis)

In Tables 4 and 5, we will present the main results of the coding of

students teachers’ argumentation per odd one out task, in order to gain a deeper understanding of the characteristics of their grammatical reasoning. For a better sense of perspective, these results will be compared to the reasoning of 14 year old pre-university students, who tackled these tasks during a short metaconceputal intervention (see Van Rijt, 2020, ch. 6). In both instances, the same coding scheme was used, enabling direct comparisons between these two participant types. The main results will then be explored in more depth for the student teachers, as these students are what we are primarily interested in. As can be inferred from Tables 4 and 5, students’ and student teachers’ grammatical argumentation can fit into one of four categories: (1) exclusion based on an appropriate grammatical argument, (2) exclusion based on an inappropriate grammatical argument (3) exclusion based on a non-grammatical argument (4) exclusion without argumentation. The tables show some interesting differences between both odd one out tasks and between student teachers and pre-university students. As for the former, the most striking finding is that in Task 2, student teachers seem to struggle much more in coming up with appropriate grammatical arguments compared to Task 1 (over one third of the arguments in Task 2 is an inappropriate grammatical argument, whereas this is just over one fifth in Task 1). This task also appears much harder for the pre-university students, as the percentage of appropriate grammatical arguments reduces by 50 percent from Task 1 to Task 2. As for the differences between student teachers and pre-university students, one would expect that student teachers would outperform pre-university students overall. However, this is not the case. Pre-university students seem to perform better at Task 1 than the student teachers, with a larger percentage of arguments being appropriate. In addition, the student teachers tend to provide much more arguments that are not about grammar than the pre-university students. As for Task 2, student teachers seem to do better overall, with a larger percentage of arguments being appropriate and a lower percentage of arguments being non-grammatical. The amount of inappropriate grammatical arguments is comparable, constituting roughly one third of the arguments in both populations.

These results provide some indication of how the tasks have been handled by both types of students. As we are primarily interested in student teachers, we will focus exclusively on their argumentation from now on. Below, we will explore what kinds of appropriate, inappropriate and non-grammatical arguments they predominantly used, and what else stands out in their responses to these odd one out tasks. Tables 6 and

Table 4

Coding scheme for odd one out Task 1 (cf. Table 2), comparing pre-university students’ grammatical argumentation (ps, N = 123 arguments) to student teachers’ grammatical arguments (st, N = 154 arguments).

| Axial codes | Definition of axial codes | Typical example from the data (ps/st) | N ps (%) | N st (%) |
|-------------|---------------------------|-------------------------------------|----------|----------|
| Exclusion based on appropriate grammatical argumentation | Response in which odd one out is chosen based on a grammatical argument that is true and appropriate. | PS: ‘To grow’ is the odd one out, because you can add a direct object to the other two verbs. | 78 (63.4) | 85 (55.2) |
| Exclusion based on inappropriate grammatical argument | Response in which odd one out is chosen based on a false or untrue grammatical argument. | PS: ‘To receive’ is the odd one out, because that needs a preposition. | 33 (26.8) | 35 (22.7) |
| Exclusion based on non-grammatical argument | Response in which odd one out is chosen based on a non-grammatical argument. | ST: ‘To smoke’ is the odd one out, because smoking is not healthy. | 4 (3.3) | 33 (21.4) |
| Exclusion without argumentation | Response in which an odd one out is chosen without any argumentation. | PS: ‘To smoke’ is the odd one out. ST: - | 6 (4.9) | 1 (0.6) |

TOTAL | 123 | 154 |

Note: Because students were able to provide multiple arguments per odd one out option, the total number of arguments exceeds the total number of participants.

In order to investigate whether there were differences between junior and senior student teachers in grammatical reasoning ability, we divided these student teachers into three categories, based on similarities in their teacher training program: (1) level 1 student teachers, which were formed by students in their first two years of a fulltime BEd track and the first-year students of a part-time BEd track; (2) level 2 student teachers, which were formed by third and fourth year fulltime BEd students, as well as second year parttime BEd students; (3) level 3 student teachers, which were formed by the masters’ students. We refer to this variable as student teacher level. To investigate the effect of student teacher level, we first ran a multilevel regression model controlling for the effect of educational institute in which we explored the effect of student teacher level on reasoning scores (Z-scores) for the odd one out tasks. Next, we included TGU scores to the model to see whether student teachers’ general understanding of grammar could predict their reasoning scores better than student level alone. In this process, we took student teacher level and the TGU score into account as fixed effects (i.e., we assumed that the effects of student teachers’ level and TGU scores were similar for all student teachers) and institution as a random effect (i.e., we assumed that odd one out scores varied across institutions). For a better interpretation of the TGU scores of the student teachers, we also examined differences in TGU scores between student teachers and secondary school students via independent samples T-tests. Because an initial exploration of the data showed a significant correlation between the number of words per task and the quality score of the raters, we also ran a third multilevel regression analysis in which the total number of words was incorporated into the model as an additional predictor (see 3.1). We thus used a step-up building strategy (West, Welch, & Galecki, 2007), and determined which of these three models was best by comparing the -2 log likelihood values of the extended model compared to the previous model, using a $\chi^2$ test.

3. Results

3.1. Characteristics of grammatical reasoning (qualitative analysis)

In Tables 4 and 5, we will present the main results of the coding of
Table 5
Coding scheme for odd one out Task 2 (cf. Table 2), comparing pre-university students’ grammatical argumentation (ps, N = 121 arguments) to student teachers’ grammatical arguments (st, N = 134 arguments).

| Axial codes                              | Definition of axial codes                                                                 | Typical example from the data (ps/st) | N ps (%) | N st (%) |
|------------------------------------------|-----------------------------------------------------------------------------------------|--------------------------------------|----------|----------|
| Exclusion based on appropriate grammatical argumentation | Response in which odd one out is chosen based on a grammatical argument that is true and appropriate. | PS: ‘On a tractor’ is the odd one out, because the other two are essential parts of speech of the verbs that are in the sentences. ST: ‘Grandma’ is the odd one out, because the other two are not the subject of the sentence. | 39 (32.2) | 68 (50.7) |
| Exclusion based on inappropriate grammatical argument | Response in which odd one out is chosen based on a false or untrue grammatical argument. | PS: ‘On a tractor’ is the odd one out, because the other two are objects and ‘op een tractor’ is an adverbial. ST: ‘many burnt-out politicians’ is the odd one out, because the other two are singular and ‘many burnt-out politicians’ is plural. | 41 (33.9) | 45 (33.6) |
| Exclusion based on non-grammatical argument | Response in which odd one out is chosen based on a non-grammatical argument. | PS: ‘On a tractor’ is the odd one out, because the other two are persons, but not the tractor. ST: ‘Many burnt-out politicians’ is the odd one out, because the other two are more about the farmland’ | 35 (28.9) | 21 (15.7) |
| Exclusion without argumentation | Response in which an odd one out is chosen without any argumentation. | PS: ‘Grandma’ is the odd one out. ST: - | 6 (5.0) | 0 (0.0) |
| TOTAL | | | 121 | 134 |

Note Because students were able to provide multiple arguments per odd one out option, the total number of arguments exceeds the total number of participants.

Table 6
Specific arguments characterized as appropriate and inappropriate grammatical for Task 1.

| Argument category | Specific argument types (appropriate grammatical arguments) | N | Sources of misconception / false arguments (inappropriate grammatical arguments) | N |
|-------------------|------------------------------------------------------------|---|---------------------------------------------------------------------------------|---|
| Valency-related arguments | The other verbs are not pseudotransitive | 2 | Terminological errors | 3 |
| | The other two verbs require a direct object | 4 | Conceptual misunderstanding of valency | 9 |
| | The other verbs serve out more roles / have a greater valency / are transitive | 21 | | |
| TOTAL | | 27 | | 12 |
| Spelling and/or morphophonological arguments | The first person singular form resembles the verb stem in the other two verbs | 3 | Terminological errors | 1 |
| | The verb stem ends with a consonant in the other two verbs | 4 | ‘z ekofschip (rule of thumb used in verb spelling) does or does not apply (without explanation) | 2 |
| | The other two verbs are weak/irregular verbs | 24 | Different conjugation (e.g., double vocals in the first person singular) | 4 |
| | Other arguments | 2 | There is a an x number of consonants or (a) specific letter(s) in (some part of) a verb* | 9 |
| TOTAL | | 33 | | 16 |
| Meaning-related arguments | Subject does not act / undergoes the action | 21 | – | – |
| | Other | 2 | Diverse | 5 |
| TOTAL | | 23 | | 5 |
| Other | The other verbs don’t take the auxiliary verb to have | 1 | The other two verbs are not auxiliary verbs | 2 |
| | The other verbs can occur in a progressive construction | 1 | – | – |
| TOTAL | | 2 | | 2 |
| GRAND TOTAL | | 85 | | 35 |

* Note While most of these claims were technically true, they were not considered appropriate grammatical arguments (i.e., too distant from grammar, or not befitting the level the student teachers should be able to perform at).
mostly related to (verb)spelling. If these issues had some relation with syntax, they were coded as appropriate grammatical arguments. In other cases, they were not. Some students teachers did not manage to come up with a better argument than 'X is the odd one out, because the other two have seven letters', which we did not consider a true grammatical argument. Similarly, 'X is the odd one out, because the other two contain more consonants' was not considered a sufficiently grammatical argument. On the other hand, arguments that related consonants to the verb stem, or arguments that pointed out that only one verb was a weak verb (i.e., does not change sound in the past tense) were considered grammatical. Surprisingly, 4 student teachers did not use relevant terms when making this argument – they would for example describe that a certain verb would or would not change sound in the past or perfect tense, without referring to the concepts of weak, strong or (ir)regular verbs, even though these terms are quite common in Dutch education and would be appropriate to draw upon. 23 student teachers also argued that certain verbs could be excluded based on their meaning (e.g., 'to smoke is the odd one out, because the subject undergoes the action / does not perform the action in the other verbs'). These arguments were considered grammatical, because they relate to a core issue in the syntax-semantics interface: the relationship between syntactic functions (e.g., subject, object) and semantic roles (e.g., agent, patient). Interestingly, while semantic roles such as agent are being addressed in teacher education programs, and should be known to senior student teachers at least, not one student teacher explicitly referred to such relevant terms, or to the overarching metaconcept of semantic role. This shows that some of the arguments may have been appropriate, but they could still be considered underdeveloped in a linguistic sense. Similar observations can be made for other types of appropriate arguments. Finally, 2 student teachers made a case that roken and groeien are not auxiliary verbs, contrary to krijgen. This was a difficult argument to evaluate. The most common use of all three of these verbs (in traditional grammar) is that they are autonomous verbs, not auxiliary verbs. Traditionally, krijgen is not considered an auxiliary. There is a construction called ‘semi-passive’ where ‘krijgen’ in a way acts like an auxiliary, but it lacks most characteristics of auxiliaries, such as the property that it has no participle form, and in any case ‘krijgen’ as an auxiliary is not a part of the grammatical training the student teachers received. As for the non-grammatical arguments, which have not been explored further in Table 6, two salient categories emerged. Several student teachers indicated (1) that ‘to smoke is the odd one out, because the other two are not bad for one’s health’ (N = 10); others predominantly pointed out (2) that smoking is a deliberate choice (N = 12).

A few things stand out from Table 7 (related to Task 2). First, a frequently used appropriate argument was that ‘grandma’ was the odd one out, because the other two are not subjects. Of course, this is true. However, the same could be said for the other options as well (i.e., ‘many burnt-out politicians’ is the odd one out, because the other two are not direct objects; ‘on a tractor is the odd one out, because the other two are not adverbials’), so it could be argued that while the argument is valid in itself, it is not a strong argument for singling out ‘grandma’, as the argument circumvents the requirement that other two options should be linked by some shared (positive) characteristic. A second aspect to notice is that student teachers wrongly classified phrases or parts of speech several times, even in quite simple grammatical categories that they should have mastered even before entering teacher education (e.g., ‘Grandma’ is the odd one out, because the other two are direct objects’, meaning that they perceive an obvious adverbial as a direct object, or, similarly, ‘Grandma is the odd one out, because the other two are adverbials’, meaning that they perceive an unmistakable direct object as an adverbial). Specific confusion also related to the type of adverbial that ‘on a tractor’ might be: two of the student teachers for instance believed it was a modal adverbial, which is indicative of a poor understanding of the metaconcept of modality. It might also raise concern that three student teachers had trouble making basic assumptions about phrases, showing confusion about the distinction between

Table 7

| Argument category       | Specific argument types (appropriate grammatical arguments) | N | Sources of misconception / false arguments (inappropriate grammatical arguments) | N |
|-------------------------|------------------------------------------------------------|---|--------------------------------------------------------------------------------|---|
| Syntactic functions     | ‘Many burnt-out politicians’ is the only direct object with two modifications | 1 | ‘Grandma’ is the odd one out, because the other two are adverbials / an adverb     | 2 |
| and phrases             | ‘Many burnt-out politicians’ is the odd one out, because the others do not contain adjectives and adverbs | 2 | ‘Grandma’ is the odd one out, because the other two are direct objects           | 7 |
|                         | ‘On a tractor’ is the odd one out, because the other phrases could both serve as direct object and subject (since they are noun phrases) | 3 | General confusion over the nature of phrases                                     | 3 |
|                         | ‘Grandma’ is the odd one out, because the other phrases do not start with a preposition | 4 | ‘Many burnt-out politicians’ is the odd one out, because the other two are singular | 4 |
|                         | ‘On a tractor’ is a prepositional object / modal adverbial / manner adverbial / adverb | 5 |                                                                                  |   |
|                         | ‘Grandma’ is the odd one out, because the other capitalized constituents are not the subject of the sentence | 28|                                                                                  |   |
|                         | ‘On a tractor’ is the odd one out, because this is the only truly optional phrase | 18| ‘X is the odd one out, because the other two phrases are or are not mandatory parts of each sentence | 2 |
| Optionality             | ‘On a tractor’ is the odd one out, because the other phrases are summoned by (the valency of) the verbs | 6 |                                                                                  |   |
| Valency-related         | ‘On a tractor’ is the odd one out, because the other phrases are summoned by (the valency of) the verbs | 6 |                                                                                  |   |
| Other                   | Other                                                      | 3 |                                                                                  |   |
| GRAND TOTAL             |                                                             | 68|                                                                                  | 45|
parts of speech and phrases, about the general nature of phrases (e.g. ‘the other two do not have a subject within their phrase’) or about which phrases can be either singular or plural (some student teachers seemed to believe that ‘on a tractor’ was singular, although the concept of number of course does not apply to prepositional phrases). Third, many student teachers (N = 20) also argued that ‘on a tractor’ was the odd one out because as an adverbial, it is the only truly optional phrase, contrary to subjects and objects. While this relates to a basic distinction between adverbials and more syntactically prominent phrases such as objects, the implication of this claim is that objects or subjects are always mandatory. This is, however, not the case (cf. Broekhuis, Corver, & Vos, 2015; Perini, 2015), and at least two of the student teachers did not seem to be aware of this. Therefore, the basis of the claim holds some merit, but the resulting implication raises some problems for students teachers’ grammatical reasoning. A fourth aspect to note is that arguments pertaining to the number of words as a core difference between the three phrases were not considered sufficiently grammatical, and where thus characterized as inappropriate grammatical arguments. Finally, several arguments seemed unrelated to grammar. 18 arguments commented on the fact that two phrases related to humans (without relating this to syntax in any way) and one to an object (‘the tractor’). One student teacher attempted to relate two of the phrases, by stating that ‘many burnt-out politicians was the odd one out, because the other two are more about the farmland’. This student teacher thus seemed incapable of coming up with a truly grammatical argument, believing that grandmothers baking cakes and farmers driving on tractors are conceptually linked.

### 3.2. Quantitative analyses

In what follows, we will present the quantitative analyses related to student teachers’ grammatical reasoning. We will start by providing some descriptive statistics and a few basic analyses (3.2.1), continue with a section on TGU scores (3.2.2) and student teachers’ reflections on the odd one out tasks (3.2.3). In section 3.2.4 we will present the multilevel multiple regression modelling.

#### 3.2.1. Descriptive statistics and data exploration

Table 8 provides some descriptive statistics related to the odd one out tasks. The data from Table 8 suggests that there may be significant differences between the number of words per task. An independent samples T-test confirmed that student teachers overall have used significantly more words for Task 1 compared to Task 2 (t(214) = 2.55, p = .011). In addition, we found significant positive correlations between the number of words and the reasoning score: r(106) = .28, p = .004.

### Table 8

| Student teacher level | Total N | M score task 1 (SD) | M score task 2 (SD) | M overall reasoning score (SD) | Mean Z score (SD) | M TGU score (SD) | Mean N of words Task 1 (SD) | Mean N of words Task 2 (SD) | Mean N of words overall |
|-----------------------|---------|---------------------|---------------------|-------------------------------|-------------------|-------------------|-----------------------------|-----------------------------|--------------------------|
| 1                     | 56      | 4.07 (1.99)         | 4.85 (1.94)         | 4.42 (1.52)                  | -0.13 (1.04)      | 12.37 (3.85)     | 20.41 (17.23)              | 13.05 (12.28)              | 16.73 (12.30)             |
| 2                     | 41      | 4.37 (1.86)         | 5.09 (2.00)         | 4.72 (1.52)                  | 0.08 (1.04)       | 12.03 (3.69)     | 15.98 (15.86)              | 13.29 (16.99)              | 14.63 (14.05)             |
| 3                     | 13      | 4.73 (1.19)         | 5.46 (0.86)         | 5.10 (0.57)                  | 0.34 (0.39)       | 14.70 (2.21)     | 20.82 (13.44)              | 15.91 (13.76)              | 18.36 (13.21)             |
| Total                 | 108     | 4.25 (1.87)         | 5.01 (1.87)         | 4.60 (1.46)                  | –                 | 12.48 (3.71)     | 18.77 (16.98)              | 13.44 (14.28)              | 16.10 (13.92)             |

Note: † This mean was based on 10 master’s students rather than 11. One master’s student left this question blank. †† Task 2 was completed by 99 student teachers (11 l1v 3, 37 l1v 2, 51 l1v 1), (attrition rate: 8.33). ††† TGU scores were obtained from 97 student teachers (attrition rate: 10.19 %). Student teacher level 1 = BEd fulltime year 1–2 and parttime year 1; level 2 = BEd fulltime year 3–4 and parttime year 2; level 3 = MEd students (parttime).

Therefore, the mean overall number of words was taken into account as a predictor at the multilevel modelling (section 3.2.4). Finally, a One Way ANOVA revealed no significant differences in TGU scores between student teachers from level 1, 2 and 3: F(2) = 2.13, p = .13.

#### 3.2.2. TGU scores

For a better interpretation of the TGU scores, we have compared them to TGU scores from 120 pre-university students from a previous intervention study in secondary education (Van Rijt, 2020 ch. 6). Of these 120 pre-university students, 60 received the same TGU version as the student teachers prior to a short metaconceptual intervention, and another 60 received the same TGU version after such an intervention. Table 9 reports the 14 year old students’ scores prior to and after a four lesson intervention and compares them to the student teachers’ scores. (See theoretical framework for more details.)

Table 9: Comparison of TGU scores between student teachers (N = 97) and 14 year old pre-university students (N = 120).

| Category                              | N   | Mean (SD) |
|---------------------------------------|-----|-----------|
| Pre-university students pre-intervention | 60  | 9.80 (3.15) |
| Pre-university students post-intervention | 60  | 12.63 (3.67) |
| Student teachers                      | 97  | 12.48 (3.71) |

#### 3.2.3. Reflections on the odd one out tasks

Student teachers were also invited to reflect on the nature of the odd one out tasks by responding to questions on a five point Likert scale: how well they think they had performed, how difficult the tasks were, and how difficult for them (t(155) = 4.66, p < .001), but they lose their edge after the 14 year olds have received such a four lesson intervention (t(155) = .025, p = .81). Student teachers also indicated on a 10 point Likert scale that they felt the TGU was quite difficult for them (M = 8.76, SD = 1.46) and that they had done their best to do well on the test (M = 9.47; SD = 0.97).

### Table 9

| Category                              | N   | Mean (SD) |
|---------------------------------------|-----|-----------|
| Pre-university students pre-intervention | 60  | 9.80 (3.15) |
| Pre-university students post-intervention | 60  | 12.63 (3.67) |
| Student teachers                      | 97  | 12.48 (3.71) |

The figure overall seems to indicate that student teachers have tackled the odd one out tasks seriously. Their scores also indicate a fair level of confidence in their ability to tackle these odd one out tasks. Interestingly, they seem to consider such reasoning tasks quite suitable for higher level students, and much less suited for lower level students.
increased the explanatory power. M3, finally, also includes the elaborated TGU model (befitting our fourth research question), which significantly predicts reasoning scores. M2 adds TGU scores to the previous models we generated. Model 1 only included student teacher level as a predictor for reasoning scores. M3 shows that for every point more on the TGU, the mean Z score for the student teachers increases with 0.06. Similarly, for every word a student teacher writes, their mean Z score increases with 0.02.

Such tasks also appear to be fairly uncommon in teacher education; some student teachers have come across them, and others have not. They do at least not appear to be a steady part of the curriculum. Using One Way ANOVA’s, we found no significant differences for any of the variables reported in Fig. 1 between the three student teacher levels, with the exception of how much effort the student teachers had put in the task (F(3,514), p < .05). A Bonferroni posthoc analysis revealed that junior Bachelor’s students (Level 1) indicated putting in significantly more effort (M = 4.30, SD = 0.71) than more advanced Bachelor’s students (Level 2, M = 3.95, SD = 0.60).

3.2.4. Multilevel multiple regression analysis

Table 10 presents the comparisons of the three multilevel regression models we generated. Model 1 only included student teacher level as a predictor for reasoning scores. M2 adds TGU scores to the previous model (befitting our fourth research question), which significantly increased the explanatory power. M3, finally, also includes the elaborateness of student teachers’ response (i.e., the mean number of words per student teacher, averaged over both tasks), as using more words may either point to inprecise writing (which may have a negative impact on reasoning quality) or to more nuanced writing (which may have a positive impact on reasoning quality). The Table shows that this last Model, which includes student level, TGU scores and mean number of words per student teacher is the best fit and accounts for most of the variation in reasoning scores. Table 11 shows the parameter estimates, standard errors and p-values for all of these models (M1-M3).

Table 11 shows that master’s students (level 3) have been taken as the base line, and the effect of other student levels (1 and 2) has been estimated based on level 3 student scores and the intercept. Students from level 2 (senior bachelor students) on average perform -0.16 less than the master’s students, and junior bachelor students score -0.35 compared to the master’s students (Model 1). As can be inferred from Table 11, student teacher level does not seem to significantly predict reasoning scores on the odd one out tasks, as student teacher level overall is not significant (in neither of the three models). In other words: senior student teachers do not outperform junior student teachers in grammatical reasoning. TGU scores and the mean number of words students wrote do predict reasoning quality. Model 3 shows that for every point more on the TGU, the mean Z score for the students teachers’ reasoning increases with 0.06. Similarly, for every word a student teacher writes, their mean Z score increases with 0.02.

4. Conclusion and discussion

4.1. Summary of research objectives

The current study pursued three objectives. First, it aimed to gain a clearer sense of what Dutch language student teachers’ grammatical reasoning looks like. In doing so, it compared student teachers’ grammatical reasoning to pre-university students’ grammatical reasoning.

Table 10
Model comparison for multilevel regression analysis.

| Grammatical reasoning score (odd one out tasks) | -2LL | Δ-2LL | df | p-value | AIC |
|-----------------------------------------------|------|-------|----|---------|-----|
| M1                                            | 302.57 |      |    | <0.01** | 312.57 |
| M2                                            | 255.29 | 47.28 | 1  | <0.05*  | 267.29 |
| M3                                            | 249.78 | 5.51  | 1  |         | 263.78 |

Note * Indicates significance at the <.001 level; ** indicates significance at the <.05 level.

Table 11

|              | M1 | M2 | M3 |
|--------------|----|----|----|
| Intercept    | 0.266 | 0.33 | 0.40 | -0.69 | 0.33 | 0.50 | -0.88 | 0.31 | 0.31 | 0.064 |
| Student level |     |     |     | -0.35 | -0.14 | -0.11 | -0.88 | 0.47 | 0.06 |
| Student level |     |     |     | -0.16 | -0.30 | -0.34 | -0.14 | 0.32 | 0.65 |
| Student level |     |     |     | 0.63 | 0.34 | 0.75 | 0.12 | 0.33 | 0.72 |
| TGU scores   |     |     |     | 0.02** | 0.02 | 0.03** | 0.06 | 0.02 | 0.03** |
| Mean N of words |     |     |     |     |     |     | 0.02 | 0.01 | 0.02** |

Note Student teacher level 1 = BEd fulltime year 1 - 2 and parttime year 1; level 2 = BEd fulltime year 3 - 4 and parttime year 2; level 3 = MEd students (parttime). Note * Indicates significance at the <.001 level; ** indicates significance at the <.05 level.
Second, the study examined whether more senior student teachers (level 2, level 3) show increased grammatical reasoning ability compared to junior student teachers (level 1), to determine the potential effects of the teacher education track on students’ grammatical reasoning competencies. Third, by means of a multilevel regression analysis, we examined what the impact of grammatical understanding (TGU) is on grammatical reasoning ability. We also investigated the impact of the elaborateness of the response (i.e., the number of words student teachers wrote in tackling the odd one out tasks) on reasoning quality.

4.2. Interpretation of main results

As hypothesized, student teachers seem to struggle with grammatical reasoning tasks overall. While there are some differences between the two tasks, just over half of student teachers’ total arguments (53 %) can be characterized as appropriate (i.e., correct and befitting of the task), whereas in over a quarter of cases their arguments were inappropriate (28.2 %). Almost one fifth of their arguments (18.6 %) does not even relate to grammar. We argue that resorting to non-grammatical arguments when specifically asked to provide a grammatical argument is indicative of insufficiently developed grammar skills and knowledge. When compared to 14 year old pre-university students, it seems that student teachers do not perform markedly better. Pre-university students’ grammatical arguments are appropriate almost half of the time (47.8 %), they are inappropriate in almost a third of all cases (30.4 %) and they pertain to non-grammatical issues 18.6 % of the time. It would seem that one of the odd one out tasks was handled better by the 14-year olds (cf. Table 4), whereas the other task (cf. Table 5) was handled better by the student teachers, resulting in slightly more favorable averages for the student teachers overall. However, recall that we hypothesized that student teachers would manage to outperform the 14 year old pre-university students on all accounts. This hypothesis can therefore not be confirmed. Likewise, we expected that student teachers would be able to outperform pre-university students on the Test for Grammatical Understanding (TGU). Surprisingly, while the student teachers did manage to outscore their secondary school counterparts before the latter had been enrolled in a brief metaconceptual intervention (see Table 9), they did no better than the secondary school students after this short intervention. While this might suggest that short metaconceptual interventions can be powerful in enhancing learners’ grammatical understanding (Van Rijt, 2020), it also points to potential shortcomings in teacher education programs as far as grammatical knowledge is concerned. We will discuss this issue in section 4.4. A potential explanation of the fact that student teachers do not perform markedly better in grammatical reasoning than pre-university students is that the former in most cases did not experience pre-university education, but higher vocational education instead. As pre-university education emphasizes abstract and critical thinking to a larger extent than higher vocational education, it might be that this distinction in student type accounts for reasoning differences later on. On the other hand, all of these student teachers will end up with at least a second degree teaching license, meaning that they will be permitted to teach 14 year old pre-university students. It might be expected, therefore, that regardless of differences in prior education, teacher education should be able to enhance students teachers’ grammatical reasoning level in such a way that student teachers can confidently outreason secondary school students. This does not appear to be the case at present. Future studies might therefore investigate the role of student teachers’ prior education. When looking at the students teachers’ argumentation in more detail, a few things stand out. First, student teachers seem to feel more secure about their spelling related knowledge than about their syntactic knowledge, given the great amount of arguments related to spelling and morpho-phonological matters in Task 1. In fact, most of the appropriate arguments were from this category. Many student teachers thus seem to be able to access this knowledge more easily than syntactic knowledge in their argumentation. Second, several of them seem capable of reasoning based on metaconcepts such as valency (although this is mainly observed in Task 1, in which valency is an obvious syntactic metaconcept to use). Fewer student teachers show signs of an incomplete understanding of this metaconcept: for every student teacher that misunderstands (aspects of) valency, three student teachers show no signs of misunderstanding this metaconcept. While this is a fair ratio, all student teachers should be able to adequately reflect on crucial metaconcepts such as valency. The odd one out tasks were initially designed for 14 year olds and therefore pertained to very basic syntactic categories (verbs, subjects, direct objects, adverbials). Given the large percentage of inappropriate responses from student teachers, it seems that many of them still struggle with such basic notions, which is cause for some concern. This concern increases if one takes into account that the student teachers have expressed that they have done their best on performing well, and that they feel fairly confident about their performances (cf. Fig. 1). This finding aligns with previous research, which revealed that university students and student teachers possess only limited Knowledge about Language (Kal) – (cf. Alderson & Cajkler & Hislam, 2002; Hudson, 2013; Van Rijt, Wijnands et al., 2019), which they tend to overestimate (Sangster et al., 2013). Surprisingly, senior student teachers do not seem to significantly outperform junior student teachers, as was shown in the multilevel modelling. This indicates that student teachers hardly seem to develop their grammatical reasoning skills in the course of their teacher training program. It should however be noted that there may be differences between level 3 students (following an MEd program) and the BEd students (level 1 and 2), which have not come up in the multilevel modelling as a result of an underpowered sample of MEd students (recall that only 11 MEd students participated in the study). The variable of student teacher level should thus be interpreted with some caution, although it would at least seem that senior bachelor students do not outperform junior bachelor students. This can also be said for their level of understanding, as TGU scores are not significantly different across student teacher levels. Two factors can significantly predict students teachers’ grammatical reasoning quality. First, their TGU scores, which indicates that the greater a student teacher’s understanding of the subject matter is, the better they are able to perform on grammatical reasoning tasks. Second, the number of words they wrote to explain their argument. Generally, the more words, the better their reasoning is scored. This finding can easily be explained by assuming that student teachers with a greater understanding are more likely to express a nuanced argument, or, from the opposite perspective: if student teachers hardly understand grammar, they are unlikely to need many words in their reasoning. This finding might inform teacher educators who are employing odd one out tasks, as it seems relevant to determine a word minimum needed to successfully complete specific tasks.

---

4 One might wonder whether there is a threshold level of grammatical knowledge that student teachers should possess, and if so, what that level might be. It makes sense to expect that student teachers should be able to outperform secondary school students on tests as the TGU at all times, but what a reasonable minimum score for teachers and student teachers is, still needs to be empirically established. Informal tests among teachers with linguistic expertise show that for most of these teachers, a score of 18 or 19 point is feasible, which seems like a sufficient score in for teachers to effectively teach grammar.
reasoning tasks. Very limited student teacher responses in terms of word count could thus be an easy warning sign for teacher educators that their student teachers may not have explored their argument for exclusion deeply enough. The current study implies that education focused on Kal should place a stronger emphasis on reasoning and grammatical understanding.

4.3. Study limitations

To the best of our knowledge this study was the first to explore the grammatical reasoning ability of student teachers in odd one out tasks. Even though it has yielded some interesting results, there are also some limitations to take into account in interpreting them. First, as previously mentioned, the number of MEd students is limited, so it is difficult to assess how master’s students perform compared to bachelor’s students. Second, while the odd one out tasks have been selected with great care, the nature of these tasks might affect the outcomes – as can be observed, the two tasks differed in terms of the types of arguments they elicited, meaning that two different tasks may have given a different idea of student teachers’ grammatical reasoning ability. Future studies might explore the effect of different reasoning tasks (related to different grammatical (meta)concepts) or different reasoning forms (i.e., different task types). A third limitation to consider is that the data from a previous study (Van Rijt, 2020, ch. 6) allow for only some comparisons between secondary school students and student teachers. However, given our interest in student teachers and general space constraints, we did not flesh out this comparison fully. Instead, we refer the reader to Van Rijt (2020, ch. 6) to gain a deeper sense of the pre-university students’ data. Fourth, the method for evaluating students teachers’ reasonings in the odd one out tasks (i.e., holistic rating on a 10 point Likert scale by two raters) warrants some caution in interpreting the results. While the two raters were both experienced and agreed to a large extent about the rating of the reasonings, future research might improve this aspect, either by including more raters or by adopting a more reliable method of assessment (e.g., comparative judgement, see Verhavert, Bouwer, Donche, & De Maeyer, 2019). Finally, the current study did not take into account student teachers trained at universities. It is likely that these student teachers are more capable of reasoning about grammar, although at this point, we can only speculate about this. We leave this matter open for future research.

4.4. Practical implications

We will end this paper with some practical implications. Overall, student teachers appear willing to use odd one out tasks in their own grammar teaching, especially for higher vocational and pre-university students, which is encouraging for grammar teaching practices in itself. After all, such tasks stimulate in-depth understanding and reasoning skills, contrary to traditional parsing exercises, which are now often at the heart of grammar lessons. Given student teachers’ struggles with grammatical reasoning, however, it will not be easy for them to adequately do this. In order for student teachers to capitalize on grammatical odd one out tasks, their own ability to handle them needs to improve greatly. Since senior BEd students do not outperform junior BEd students, a first recommendation would be to spend more time and attention to grammatical and linguistic reasoning in the entire teacher education program, especially in the later phases of the track. This can either be done in the form of odd one out tasks, or any other tasks that stimulate grammatical reasoning. Simply spending more time on traditional grammar in itself seems insufficient to boost students teachers’ grammatical reasoning ability. Given that TGU scores can predict reasoning quality, teacher education programs should invest in grammatical understanding more, for example by implementing linguistic metaconcepts (such as semantic roles, which the students did not use at all), and incorporate reasoning tasks in their grammar teaching throughout the curriculum. This way, student teachers can experience first-hand how their grammatical understanding can impact on their ability to tackle grammatical problems (for examples, see Van Rijt, 2020, ch. 6). Student teachers should then be taught how their own content knowledge and reasoning abilities can be pedagogically translated, as pedagogical subject knowledge is vital for effective grammar teaching (Myhill et al., 2013). Teacher educators have a critical role in guiding student teachers’ grammatical reasoning, and they need to show students what nuanced grammatical reasoning (within and outside of the classroom) looks like. This way, student teachers will be more prepared to teach thinking skills in their own grammar lessons. Finally, while the current study has been conducted in the Netherlands, its relevance is much broader. In other educational contexts too, for example in those that embrace more contextualized forms of grammar teaching (e.g., Australia (Troyan, Harman, & Zhang, 2020), New Zealand (Gordon, 2005) or the United States (Accurso & Gebhard, 2020), being able to think about grammar and developing an adequate understanding of the workings of grammar is an essential goal for both in-service teachers and student teachers. This study has shown that odd one out tasks can be a good way to examine students teachers’ grammatical understanding and reasoning ability, as such tasks foreground grammatical argumentation and allow for thinking from multiple perspectives. Odd one out tasks can thus constitute a valuable addition to existing grammar teaching practices.

Acknowledgements

We wish to express our gratitude to four (former) master’s students of Fontys University of Applied Sciences (Sittard) for their help in gathering and analyzing some of the data: Inge Bastings, Anne Landman, Diana Schulteis and Robin Velraeds. We are also grateful to the anonymous reviewers of this paper for their valuable suggestions.

Appendix A. – version of the Test for Grammatical Understanding (TGU) used in the current study (translated from Dutch by the authors*)

The alternatives are ranked from full understanding (A), to partial understanding (B), blind concept use (C) and no concept use (D).

* Note Due to translations, several of the questions make less sense in English. These questions have been marked with an asterisk (‘*’), and some Dutch translations have been given in those cases to facilitate understanding. All questions have been carefully composed by two Dutch linguists (both authors of the present paper) and the TGU has been carefully pretested.
J.H.M. van Rijt et al.

Studies in Educational Evaluation 70 (2021) 101007

12

(12) Why can you say ‘a book’ in the sentence ‘Ian reads a book’, whereas you cannot say the in ‘Ian burns a book’?
A. The verb ‘to read’ selects a mandatory direct object.
B. The direct object ‘a book’ can sometimes be left out, and sometimes not.
C. The past participle ‘seen’ sounds similar to the fake past verb ‘ran’.
D. You can’t say ‘what you burn up without saying what you are burning up.

(13) Some sentences have more than one meaning, such as: ‘Stan has seen the men with the violin.’ What causes the difference in meaning?
A. The phrase ‘to make’ can get it’s meaning from the verb or from the connection with the noun.
B. You can place both ‘the men’ and the men with the violin before the fake verb ‘saw’.
C. The phrase ‘will be’ is ambiguous in this sentence: it can either mean ‘together with’ or ‘with the help of’.
D. It depends on who hold the violins in reality, Louise or the men.

(14) The following sentence is in the school paper: ‘Studenten maken een nieuw schoolblad.’
B. dishwashers clean (or calculate on a new blackboard)
C. the main verb ‘to clean’ is not a direct object.
D. the main verb ‘to clean’ is not a direct object.
E. the verb ‘to clean’ is not a direct object.
F. the verb ‘to clean’ is not a direct object.

(15) Why can you say ‘I want to get a haircut’?
A. ‘I want to get a haircut’ gives you a chance to talk about your hair.
B. ‘I want to get a haircut’ gives you a chance to talk about your hair.
C. ‘I want to get a haircut’ gives you a chance to talk about your hair.
D. ‘I want to get a haircut’ gives you a chance to talk about your hair.

(16) The following two sentences appear to be structurally the same way:
A. ‘I don’t know what he is doing’
B. ‘I don’t know what he is doing’
C. ‘I don’t know what he is doing’
D. ‘I don’t know what he is doing’

(17) Why can you say ‘I want to get a haircut’?
A. Because the verb ‘to clean’ does not select a question when it is its direct object.
B. Because the verb ‘to clean’ cannot select the direct object of the man who wants it.
C. Because the verb ‘to clean’ cannot select the direct object of the man who wants it.
D. Because the verb ‘to clean’ cannot select the direct object of the man who wants it.

(18) The sentence ‘He was the best of all the men’s team’ is ambiguous. Why is that?
A. ‘He was the best of all the men’s team’ always sounds like a lampoon.
B. ‘He was the best of all the men’s team’ always sounds like a lampoon.
C. ‘He was the best of all the men’s team’ always sounds like a lampoon.
D. ‘He was the best of all the men’s team’ always sounds like a lampoon.

(19) What is the sentence ‘He was the best of all the men’s team’.
A. ‘He was the best of all the men’s team’.
B. ‘He was the best of all the men’s team’.
C. ‘He was the best of all the men’s team’.
D. ‘He was the best of all the men’s team’.

(20) The following two sentences are the same in meaning:
A. ‘I am waiting for better weather’.
B. ‘I am waiting for better weather’.
C. ‘I am waiting for better weather’.
D. ‘I am waiting for better weather’.

(21) The sentence ‘I am waiting for better weather’ is grammatical, but it is not grammatical when it is used figuratively.
A. In ‘I am waiting for better weather’, the proposition that has no meaning of place.
B. In ‘I am waiting for better weather’, the proposition that has no meaning of place.
C. In ‘I am waiting for better weather’, the proposition that has no meaning of place.
D. In ‘I am waiting for better weather’, the proposition that has no meaning of place.

(22) The sentence ‘I am waiting for better weather’ is grammatical, but it is not grammatical when it is used figuratively.
A. In ‘I am waiting for better weather’, the proposition that has no meaning of place.
B. In ‘I am waiting for better weather’, the proposition that has no meaning of place.
C. In ‘I am waiting for better weather’, the proposition that has no meaning of place.
D. In ‘I am waiting for better weather’, the proposition that has no meaning of place.
References
Accurso, K., & Gebhard, M. (2020). SFL praxis in U.S. teacher education: A critical literature review. Language and Education: Advance online publication, Alderson, J., & Hudson, R. (2013). The metalinguistic knowledge of undergraduate students of English language or linguistics. Language Awareness, 22(4), 320–337.
Andrews, R. (2005). Knowledge about the teaching of [sentence] grammar: The state of play. English Teaching Practice & Critique, 4(3), 69–76.
Andrews, R. (2010). Teaching sentence-level grammar for writing: The evidence so far. In T. Locke (Ed.), Beyond the grammar wars. A resource for teachers and students on developing language knowledge in the English/literacy classroom (pp. 91–108). New York: Routledge.
Andrews, S. (1997). Metalinguistic awareness and teacher explanation. Language Awareness, 6(2–3), 147–161.
Awramiuk, E., & Szymańska-Vaugh, (2012). A qualitative study into the effect of contextualised grammar teaching on student writing. Research Papers in Education, 29(3), 130.
Baumberger, C., Beisbart, C., & Brun, G. (2016). What is understanding? An overview of the concept of understanding. Language and Linguistics Compass, 10(3), 1–24.
Bakken, H., Cover, N., & Vos, R. (2015). The syntax of Dutch: Verbs and verb phrases. Part I. Amsterdam, The Netherlands: Amsterdam University Press.
Cajker, W., & Hilsam, J. (2002). Trainee teachers’ grammatical knowledge: The tension between public expectation and individual competence. Language Awareness, 11(3), 161–177.
Caling-Jageman, R., & Cumming, G. (2018). The new statistics for better science: Ask how much, how uncertain, and what else is known. PsyArXiv Preprints.
Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6(4), 284–290.
Corbin, J., & Strauss, A. (2015). Basics of qualitative research: Techniques and procedures for developing grounded theory. Thousand Oaks: Sage.
Curriculum.nl (2019). Conceptvoorzieningen leerovich Nederlands, 2019.
De Rijt, J. (2009). The epistemic value of understanding.
De Regt, H. (2009). The epistemic value of understanding.
Describing verb valency. Practical and theoretical issues
Disney, M. (2000). Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1, 30–46.
Gordon, E. (2005). Grammar in New Zealand schools: Two case studies. Language and Education, 17(5), 355–370.
Hudson, R. (2004). Why education needs linguistics (and vice versa). Journal of Linguistics, 40(1), 105–130.
Honda, M., & Graus, J. (2018).
Meestersinga, T.B., & Ravenscroft, C. (2013). Het schoolexamen Nederlands havo/vwo in kaart. Levende Talen Tijschrift, 14(2), 13–19.
McGraw, K., & Wong, S. (1996). Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1, 30–46.
Mcintire, I., & Wijnands, A. (2012). Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1, 30–46.
McIntyre, I., & Wijnands, A. (2012). Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1, 30–46.
Meijerink, H., et al. (2008). Over de drempels met taal en rekenen - Hoofdstuktwo: De Expertgroep Doelopende Leerlijnen Taal en Rekenen. Enschede.
Myhill, D. (2000). Misconceptions and difficulties in the acquisition of metalinguistic knowledge. Language and Education, 14(3), 151–163.
Myhill, D. (2003). Principled understanding? Teaching the active and passive voice. Language and Education, 17(5), 355–370.
Myhill, D. (2018). Grammar as a meaning-making resource for improving writing. L1 Educational Studies in Language and Literature, 18, 1–21.
Myhill, D., Jones, S., Lines, H., & Watson, A. (2012). Re-thinking grammar: The impact of embedded grammar teaching on students’ writing and students’ metalinguistic understanding. Research Papers in Education, 27(2), 151–163.
Myhill, D., Jones, S., & Watson, A. (2013). Grammar matters: How teachers’ grammatical knowledge impacts on the teaching of writing. Teaching and Teacher Education, 36, 77–91.
Onderwijs in Cijfers. (2019). Lerarenopleidingen. Consulted via https://www.onderwijs cijfers.nl/kengetallen/onderwijs-algemeen/personeel/lerarenopleidingen on December 29, 2020.
Perini, M. (2015). Describing verb valency. Practical and theoretical issues. New York: Springer.
Ratty, K., Awramiuk, E., & Fontich, X. (2019). What is grammar in L1 education today? L1 Educational Studies in Language and Literature, 19, 1–8.
Sangster, P., Anderson, C., & O’Hara, P. (2013). Perceived and actual levels of knowledge about language amongst primary and secondary student teachers: do they know what they think they know? Language Awareness, 22(4), 293–319.
Troyan, F., Harman, R., & Zhang, X. (2020). Critical SFL praxis in teacher education: Insights from Australian SFL scholars. Language and Education. Advance online publication.
Van der Aalvoort, M. (2016). Vensters op vakontwikkeling. De betewe invorder van taalkunde in het examenprogramma Nederlands havo/vwo (1988-2008). Nijmegen: Radboud Universiteit. Doctoral dissertation.
Van Gelderen, A. (2010). Does explicit teaching of grammar help students to become better writers? Insights from empirical research. In T. Locke (Ed.), Beyond the grammar wars. A resource for teachers and students on developing language knowledge in the English/literacy classroom (pp. 109–128). New York, NY: Routledge.
Van Rijt, J. (2020). Understanding grammar. The impact of linguistic metaconcepts on L1 grammar education. Doctoral dissertation. Nijmegen: Radboud University. https://doi.org/10.13140/RG.2.2.31243.62249.
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/10.1080/09658416.2017.1416552.
Van Rijt, J., De Swart, P., Wijnands, A., & Coppen, P.-A. (2019). When students tackle grammatical problems: Exploring linguistic reasoning with linguistic metaconcepts. Language and Education, 32, 78–93. https://doi.org/10.1080/09658416.2019.1630037.
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/10.17229/LESSL-2019-19-02-03.
Van Rijt, J., Wijnands, A., & Coppen, P.-A. (2020a). How secondary school students may benefit from linguistic metaconcepts to reason about L1 grammatical problems. Language and Education, 34(3), 231–248. https://doi.org/10.1080/09658416.2020.1784526.
Van Rijt, J., Wijnands, A., & Coppen, P.-A. (2020b). Investigating Dutch teachers’ beliefs on working with linguistic metaconcepts to improve students’ L1 grammatical understanding. Research Papers in Education, 1–29. https://doi.org/10.1080/09578219.2019.1690562.
Van Rijt, J., Wijnands, A., & Coppen, P.-A. (2020c). Dutch teachers’ beliefs on working with linguistic metaconcepts to improve students’ L1 grammatical understanding. Research Papers in Education, 1–29. https://doi.org/10.1080/09578219.2020.1784526.
Verhavert, S., Bouwer, R., Donche, V., & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. Assessment in Education Principles Policy and Practice, 26(5), 541–562.
Watson, A. (2012). Navigating ‘the pit of doom’: Affective responses to teaching ‘English’. English in Education, 46(1), 22–37.
Watson, A., & Newman, R. (2017). Talking grammatically: L1 adolescent metalinguistic reflection on writing. Language Awareness, 26(4), 381–398.
Welling, J. (2000). Educational research: Contemporary issues and practical approaches. London: Continuum.
West, B., Welch, K., & Galecki, A. (2007). Linear mixed models: A practical guide using statistical software. New York, NY: Chapman Hall/CRC.