POLITENESS WITHOUT PAYOFF? EVALUATING THE EFFECT OF A SHORT PRAGMATICS INTERVENTION ON L1 REQUEST WRITING

AFRA KLARENBEEK[1] AND JIMMY VAN RIJT[2]

1. Tilburg Center of the Learning Sciences, Tilburg University, the Netherlands
2. Department of Languages, Literature and Communication, Utrecht University, the Netherlands

Abstract
This study explores the introduction of pragmatics concepts—speech acts, conversational implicature, and politeness—into L1 language education and its impact on students' writing. Despite the well-documented benefits of pragmatics instruction in L2 contexts, its application in L1 settings remains underresearched. Using a quasi-experimental research design with switching replications, involving 241 Dutch pre-university students, the current study examined whether a short intervention in which core pragmatics concepts were introduced would benefit the quality of students' written requests, their self-efficacy in writing them and the associated perceived mental effort. Contrary to expectations, the intervention did not manage to secure any improvements on these variables. While this result may be due to the nature of the intervention in terms of its design principles or duration, we also found evidence that the number of words students wrote during the tests significantly declined over time, revealing a decrease in motivation for the writing tasks. This appears to be a major contributing factor to the results. Based on these results, the paper discusses meaningful ways forward for future research on this topic, emphasizing the need for a better understanding of student learning through improved fidelity measures and as well as following strategies to maintain student engagement during repeated testing.

Keywords: pragmatics, politeness theory, writing, self-efficacy, linguistics education, secondary education, comparative judgment

Klarenbeek, A., & Van Rijt, J. (2025). Politeness without payoff? Evaluating the effect of a short pragmatics intervention on L1 request writing. L1-Educational Studies in Language and Literature, 25, 1-26. https://doi.org/10.21248/l1esll.2025.25.1.883
Corresponding author: Jimmy van Rijt, Department of Languages, Literature and Communication, Utrecht University, Trans 10, 3512 JK Utrecht, email: j.h.m.vanrijt@uu.nl

1. INTRODUCTION

One of L1 school subject's principal goals is to teach students how to communicate effectively, both in writing and in oral interaction with other language users (Green & Olaf-Erixon, 2020; Sawyer & Van de Ven, 2007). Part of teaching students how to communicate effectively ideally entails providing students not just with technical language skills (e.g., writing or presentation skills) but also with a deeper understanding of how communication works. In other words: students would also benefit from obtaining knowledge about communication and communicative processes such as knowledge about speech acts (e.g., Austin, 1962; Searle, 1969), conversational implicature (e.g., Cummings, 2005) and politeness theory (e.g., Brown & Levinson, 1987). This way, students' language skills might be underpinned by an understanding of critical communicative mechanisms and patterns, which arguably makes their language skills more conscious. This unification of language skills and knowledge about language can best be understood from a perspective of 'powerful knowledge' (Young, 2013), where it is believed that learners are entitled to understand how language and communication work, rather than seeing them as merely a tool (Sheehan et al., 2021; Van Rijt & Coppen, 2021). This also entails that language education should move beyond 'everyday knowledge' about language, but that it should rather offer forms of disciplinary knowledge (Young & Muller, 2013), that is powerful in the sense that it offers students new ways of thinking about the world, of which language is a fundamental part.

To introduce students to such powerful knowledge about communicative processes, language education could draw from the related scientific discipline of linguistics, and more specifically 'pragmatics', which examines the use of natural language in everyday communication. To do so adequately, the academic knowledge offered by pragmatics would have to be recontextualized (a concept attributed to Bernstein, 1999) to align with educational objectives (Hordern, 2021). While pragmatics could potentially be enriching to language education, in L1 contexts, explicit insights from this academic discipline, such as the aforementioned speech acts, conversational implicature and politeness theory, only seldom find their way into language curricula (with some exceptions of course, see Clark's (2016) account). This is quite surprising, since pragmatic knowledge would align well with the communicative skills-oriented goals of language education, whilst at the same time contributing to the frequently heard desire of introducing more content knowledge into the language curriculum, including linguistic knowledge (e.g., Denham & Lobeck, 2010; Denham, 2020; Dera et al., 2023; Hudson, 2004; Van Rijt & Coppen, 2017).

In L2 or foreign language education, things are rather different. In those contexts, pragmatics education has been studied fairly extensively, especially compared to the L1 situation. A recent meta-analysis (Ren et al., 2023) shows that pragmatics instruction in L2 and foreign language studies has a large, positive effect on the development of students' pragmatic competence in the target language, both when the instruction was explicit and implicit. The meta-analysis also shows that longer

treatments tend to yield larger effect sizes compared to shorter ones, where it was also found that students appear to be most susceptible to pragmatics instruction during their high school years (yielding the largest mean effect size of all the educational levels taken into account in the review).

Given these promising results in non-L1 contexts, it might be expected that L1 language education could also benefit from the introduction of pragmatic concepts, not only providing students with an understanding of the workings of their L1, but also potentially transferring to their writing in situations that are pragmatically challenging (Taguchi et al., 2022), e.g., making polite requests over email—a genre that students seem to experience real difficulty with (Biesenbach-Lucas, 2007; Economidou-Kogetsidis, 2011, 2016). However, to the best of our knowledge, no empirical studies to date have examined the impact of a pragmatics module on students' writing in L1 settings. This study attempts to empirically explore the added value of the teaching of core pragmatic concepts in such contexts. It tries to answer the following research question:

> To what extent does a short intervention targeting explicit pragmatics concepts (speech acts, conversational implicature and politeness) in an L1 context benefit pre-university students' written requests?

Before explaining the methodology, we will first clarify (1) the pragmatic concepts at the heart of our study; (2) design principles underpinning the intervention that are derived from other work in which linguistic concepts were introduced in L1 education; (3) the Dutch educational context in which this study is set.

1.1 Key pragmatic concepts

Some of the key concepts in scientific pragmatics include speech acts, conversational implicature and politeness theories. The intervention that we report on, has attempted to introduce students to these concepts in a coherent way, as they can best be understood in relation to one another.

Speech acts are described by Austin (1962) as the actions performed in saying something—they are thus actions that come into being by using language. Speech acts are in that sense quite distinct from 'regular' acts. For example, when an individual utters the words 'Of course' in response to a request bring a book for someone, they have in fact performed the speech act of *making a promise*. The promise has come into the world simply by using language; by contrast, 'regular' acts, such as *baking a cake*, do not simply come into being by uttering the words 'I am baking a cake' (Houtkoop & Koole, 2008, p. 20). Speech acts thus play a pivotal role in everyday communication, as it matters whether we have performed a certain speech act or not (e.g., did you make a promise when uttering the words 'Of course' or could you later plausibly deny that you did?). Traditionally, speech acts are analyzed from three perspectives: the words that are actually said ("Of course", also called the

locution), the underlying speech act ("making a promise", also referred to as *illocution*) and the result of the words (opening the window or not, also called *perlocution*).

Conversational implicatures are forms of indirect speech acts, where the hearer has to infer what is meant by a certain utterance (Cummings, 2005; Grice, 1975). Such indirect speech acts are context-dependent and individuals infer the intended meaning because they assume that contributions to a conversation are generally relevant to the direction of the conversation. Individuals are taken to cooperate to assure that a shared direction in the conversation occurs (the so called 'cooperation principle'), which is underpinned by so called 'maxims': general principles that individuals adhere to in communicative contexts, namely the maxim of quantity, of quality, of manner and of relation (for details, see Grice, 1975). Based on this general principle of cooperation, individuals can infer that the utterance 'My dog is ill' might be understood as declining an invitation to go to the movies tonight. While there have been some theoretical advances since Grice's seminal work (e.g., Horn, 2007; Levinson, 2000), for secondary education, it makes sense to focus on the initial work, because in addition to being foundational to this later work, it is fairly easy to comprehend, and students are reported to have some fun working with Gricean maxims (Clark, 2016).

Politeness is, in Brown's words, 'crucial to the construction and maintenance of social relationships, [going] to the very heart of social life and interaction; indeed it is probably a precondition for human cooperation in general' (2015, p. 326). Given the importance attributed to politeness in communication, it makes sense for education to pay attention to its underlying communicative mechanisms. This will require a different perspective towards politeness than that usually taken by the general public. For the layman, politeness is often seen as a concept 'designating 'proper' social conduct, rules for speech and behavior stemming generally from high-status individuals or groups', rules that are 'often formulated in etiquette books' (ibid.).

The linguistic perspective on politeness is different, however, where politeness is either seen as either a speaker's adherence to politeness maxims (e.g., adjust your contribution in a way that the communicative context demands), such as the maxims of quality, quantity, relevance and manner, cf. Grice, 1975) or as a form of 'face management'. This latter perspective towards politeness has been made influential by Brown and Levinson (1987), who, following Goffman's earlier work (1967), have defined the notion of face as 'an individual's publicly manifest self-esteem', proposing that language users have two opposite face requirements: each individual has a desire to be appreciated and valued by others ('positive face') as well as a desire to not offend others, thus protecting their face ('negative face'). This stems from a desire to have the liberty to make one's own autonomous decisions. Language users subsequently develop strategies to accommodate their own and other language users 'face needs', in particular when speech acts are at play which are potentially face-threatening, so called *Face Threatening Acts* (FTA's). For example, when someone

makes a request (an FTA), the speaker will adopt strategies to maximize the likelihood of that request being fulfilled. They can, for instance, accommodate the hearer's desire for positive face ('Paul, you are always so good at these things') before making the actual request ('so would you please help me with this email for a minute'), or they can minimize the impact for the hearer, hedge the request, excuse themselves or add modality to 'soften' the request (see the example in the previous sentence—*would, please, for a minute*)—see e.g., Cutting (2015) for more details on politeness strategies.

In deciding how much politeness is required in a given situation (e.g., when making a request or correcting someone), speakers will generally take contextual cues into account. Apart from potential intercultural differences, which might have a profound impact on how politeness is perceived, individuals will take at least three variables into account (Brown, 2015; Van der Wijst, 1996). These are (1) the power distance (the speaker's hierarchical relationship to the hearer, with more politeness being necessary when addressing individuals who are higher up in the hierarchy); (2) the social distance (the social distance between speakers, i.e., one would address their hair dresser differently than a close friend); and finally (3) the rate of imposition (where ' heavier requests' demand more politeness than lighter ones, as they put a stronger claim on the hearer's autonomy—asking someone to lend five euro's is a much lighter request than asking some to lend 500). As is previously touched upon, while Brown and Levinson's face keeping theory attempts to have universal applications (i.e., across cultures and languages), their work has been criticized for their claims towards universality (e.g., Culpeper, 2011; Song, 2017).

The intervention that is at the heart of this paper has explicitly targeted all of the abovementioned pragmatics concepts (see Table 1 for an overview), adopting the aforementioned notions, including face keeping, although students were informed that the face keeping theory cannot be extrapolated to all intercultural contexts (McConachy, 2013).

1.2 Design principles for the introduction of linguistic concepts

When recontextualizing concepts from academic linguistics to L1 language education, it is needed to pay specific attention to the pedagogic and didactic design principles that may best facilitate students' understanding of these concepts. While we are not aware of specific design principles for the introduction of pragmatic concepts to L1 education, we have underpinned our intervention with effective design principles from interventions that have successfully introduced students to abstract syntactic concepts (e.g., Myhill et al., 2012; Van Rijt et al., 2020) and with insights from L2 pragmatics research (e.g., Ren et al., 2023). We will briefly discuss these below.

1. Key concepts from pragmatics will be introduced inductively and implicitly

This first design principle aligns with those from studies that have attempted to stimulate students' language intuitions about the relevant concepts by means of guided induction (see Haight et al., 2007; Van Rijt et al., 2020, 2022), following Moseley et

al.'s (2005) recommendation that thinking skills ideally need an initial stage of *information gathering*, in which learners are exposed to experiences and recall information that they already recognize (see also Wijnands et al., 2021). When introducing the concept of *speech act*, this was done, for example, by first asking students how they would naturally react to a certain situation, before discussing their initial reactions in a more fine-grained way. In one such task, students were for example asked to reflect on a dialogue in which the student had allegedly lend his school book to a classmate named Finn. When the student asks if Finn would bring the book back tomorrow, he responds with 'No problem'. The next day, Finn did not bring the book back, to which the student says: "But you promised!". Finn's response to this is that "he did not use the word *promise*" so it does not count. Students are then asked to indicate whether they feel that Finn's reaction is justified. This does not only inductively and implicitly introduce speech acts, but also opens up windows for discussing associated concepts such as *locution* versus *illocution*. As illustrated by the example above, students are usually asked to think about solving a problem in the pragmatics domain during this stage.

2. After the inductive introduction, pragmatics concepts will be linked to the students' experiences explicitly, using relevant linguistic terminology

Following the inductive introduction of a key concept, an explicit link is then made (through both the teacher and the texts from the materials) between the insights that have come out of the inductive task and an explicit concept from pragmatics. This recommendation is in line with Moseley et al.'s (2005) second stage of thinking and learning, namely *building understanding*, in which the initial experiences from the information gathering stage are deepened and expanded, thus contributing to organizing ideas and concept formation. Combining implicit experiences and problem-solving from the first design principle with explicit instruction later on seems to benefit learning (see e.g., Sinha & Kapur, 2021). This principle aligns with findings from L2 pragmatics, showing that both explicit and implicit instruction might be favorable to the learning of L2 pragmatics (Ren et al., 2023).

3. Pragmatic concepts must be related to everyday communicative situations by means of practical examples and cases

A useful way of recontextualizing concepts from pragmatics for language education, is to connect these concepts to everyday communicative contexts that students recognize. Using authentic (con)texts has proved to be useful in educational material in which grammatical concepts and writing are linked (e.g., Myhill et al., 2012; Steenbakkers, 2023); quite possibly, pragmatics education could benefit from a similar approach, as pragmatic concepts intrinsically deal with actual communicative contexts. For this reason, this approach is advocated in a Dutch didactic handbook for the teaching of pragmatic concepts to secondary school students (Wegman & Van Rijt, 2023). In the material, cases were selected in co-consultation with a secondary school teacher to ensure this, and most of the examples and cases used take the students' perspective as a point of departure.

4. The material promotes linguistic reasoning

Given the context-dependent application of pragmatic concepts and the fact that pragmatic concepts can be applied to these contexts in different ways, the material has attempted to stimulate linguistic reasoning (Dielemans & Coppen, 2021; Van Rijt, 2024) for example by encouraging students to take multiple perspectives into account before answering a question, by comparing their own intuitions with those of other students (adhering to socio-cultural principles of metalinguistic activity, cf. Fontich, 2016) and by foregrounding students' reasoning over their answers alone. The teacher was asked to scaffold different reasoning processes, and to pay attention to the reasoning process in whole-class reflections afterwards. This design principle aligns with Moseley et al.'s (2005) stage of *productive thinking*, the final stage of thinking education, in which reasoning, understanding causal relationships, systematic enquiry, problem-solving and creative thinking take place.

A practical limitation for the intervention was that it needed to be fairly short: teachers had to be able to implement it within two weeks' time due to their busy schedules. While this conflicts somewhat with Ren et al.'s (2023) finding that longer treatments of pragmatics generate the largest effect sizes, this practical limitation could not be ignored. It was decided to develop four lessons of fifty minutes each (the regular amount of time for Dutch lessons), thus accommodating principles of ecological validity. Previous studies on teaching grammatical (meta)concepts have shown that such a duration can be sufficient to boost students' (morpho)syntactic awareness and reasoning skills (e.g., Dols-Koot & Van Rijt, 2019; Van Rijt et al., 2020, 2022), and therefore, it was assumed that a similar time frame would be plausible for teaching pragmatic concepts as well.

1.3 The Dutch context

The current study takes place in the Netherlands at a time in which a large national curriculum change is ongoing (SLO, 2024). At the time of writing this article, the direction of this curriculum change is that contrary to the previous curriculum, in which linguistics and literature were severely marginalized (cf. Van der Aalsvoort, 2016; Bonset & Hoogeveen, 2010; Hulshof & Van Rijt, 2020), more room will be made for linguistic and literary knowledge, and stronger links between such knowledge and language skills are aspired to (Bax et al., 2024; Dera et al., 2023; SLO, 2024). Yet pragmatics, the linguistic field that is uniquely equipped to fulfill both of these needs, remains relatively hidden in the curriculum proposals thus far—a situation that has been labeled problematic by some (e.g., Breetvelt, 2024; Wegman & Van Rijt, 2023). This article contributes to the empirical knowledge base about the introduction of pragmatics to L1 curricula, not just in the Dutch context, but certainly beyond.

The intervention that was at the heart of this study thus aligns with the new direction of the L1 curriculum, but it may be somewhat at odds with actual classroom practice, as most teachers and their students are not yet accustomed to working with topics from modern linguistics beyond school grammar (Van Rijt, 2022). In fact, only

a minority of schools seem to teach linguistics topics in the higher grades of upper secondary education 'havo/vwo bovenbouw'), even though they are allowed to—the best estimates are that just over 5 percent of secondary schools cover linguistic topics in the upper grades (ibid.). Usually, even less disciplinary linguistics is taught in the first grades, where the current study is set. When schools do teach disciplinary linguistics, pragmatics seems to be a topic that is moderately covered at the pre-university level (Van Rijt, 2022, p. 44).

2. METHOD

2.1 The intervention

As described above, the intervention consisted of four lessons (50 minutes each) that were underpinned by the theoretical design principles listed in the introduction. Each lesson centered around one main topic (see Table 1), with politeness taking centre stage in two lessons.

Table 1. Overview of pragmatics concepts covered in each of the intervention's lessons

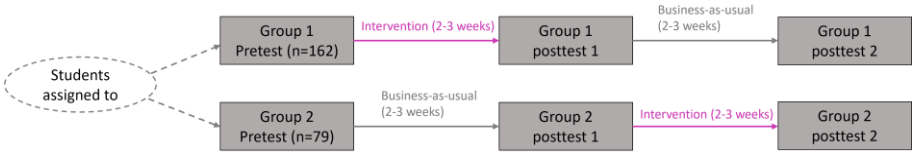
Lesson	Main topic	Relevant (sub)concepts
1	Speech acts	Performative verb, direct vs. indirect speech acts, types of speech acts, locution, illocution, perlocution
2	Cooperation principle and implicature	Gricean maximes
3	Politeness	Everyday politeness vs. linguistic politeness, Face Threatening Acts, factors related to politeness (power distance, social distance, rate of imposition)
4	Politeness	Politeness strategies (positive and negative)

To support teachers as best we could (especially given their overall unfamiliarity with teaching pragmatics), we developed all learning materials for them, including a student work booklet and a detailed teachers' manual, in which the goal of the study, its design principles and all of the tasks were explained in detail. In addition, teachers were extensively briefed orally in advance, and they were provided with theoretical explanations of the pragmatics concepts to ensure they would have access to sufficient linguistic knowledge to teach. In the spirit of ecological validity, teachers were informed that they were allowed to deviate from the material to an extent if the occasion called for it, as long as they acted in the spirit of the design principles and the learning objectives (cf. Lagemann, 2002). They were also asked to report any major deviations (see ecological validity). Teachers all taught the intervention within a period of two to three weeks (depending on their schedules), and afterwards, they all reported that students generally enjoyed the learning materials.

2.2 Research design

To answer our research question of whether teaching pragmatic concepts could translate to students' writing of requests, we adopted a quasi-experimental research design with switching replications (De Maeyer, 2021; Shadish et al., 2002). In this design, participating students are randomly assigned to one of two groups (labeled group 1 and group 2), both of which receive the intervention at different moments in time. See Figure 1 for an overview.

Figure 1. Visual overview of the switching replications design. Business-as-usual means that students received their regular education (e.g., reading, spelling, grammar); the only exception was that teachers were asked not to teach anything related to pragmatics or writing to avoid confounding the results.



The advantages of the design are manifold: in using three tests (a pretest and two posttests), the design allows for the potential detection of a 'long-term' effect (i.e., beyond an immediate posttest), as well as having the potential of ruling out a potential testing-effect, given the pretest-posttest1 setup for group 2 (without an intervention). Additionally, group 2 serves as an internal control for group 1, as the design offers the possibility of replicating an effect found in group 1 between the pretest and posttest 1. Finally, the design has important ethical advantages, as both groups receive the treatment. An a priori power calculation using G*Power (F-tests, repeated measures, within-between interaction) revealed that a total of 66 participants would be needed to be able to detect an effect size of .20 with a power of .95.

2.3 Participants

The study received ethical approval by the Research Ethics and Data Management Committee of Tilburg University under case number REDC 2022.42. The Schools, participating teachers, students and students' parents were actively asked for permission to participate in the research. If students or their parents decided not to participate in the study, the student's data was not collected.

The intervention was delivered by seven (fairly) experienced teachers from different schools that volunteered for this task (1 male). Their experience as a Dutch language teachers ranged from 4-20 years and they all actively consented to their students' data being used for the purposes of this paper anonymously. Due to ethical considerations, we were not allowed to gather more of their demographic data.

The students (N = 241) underwent the intervention during regular class hours (within the school subject of Dutch language and literature). Three students' data

were removed from the dataset because their tasks showed that they had not taken them seriously, leaving a total of 238 students, most of whom were either 14 or 15 years of age (due to ethical considerations, we only describe their age at a group level). 111 of them identified as male (46,6%); 120 identified as female (50,4%), and 10 students either identified as 'other' or preferred not to disclose this information (2,9%). Given the variability in cultural perceptions of politeness, we also gathered information on students' cultural and linguistic backgrounds. 220 students indicated having been born in the Netherlands (92,4%). The remaining 18 students came from different geographic regions. Slightly over half of them had a (North)western-European background (n = 10), the others originated from other parts of the world: The United States (n=1), Russia (n=1), Somalia (n=2), Spain (n=1), Syria (n=2) and Türkiye (n=1). Additionally, 35 students reported their mother was born outside the Netherlands (14.7%), while 23 indicated their father was foreign-born (9.7%). Furthermore, 39 students (16.4%) spoke languages other than Dutch at home, including English (n=11), Frisian (n=8), Arabic (n=3), and Russian (n=2). Some students reported speaking multiple languages at home. Chi-squared analyses indicate that there were no significant differences between the two conditions (group 1 and 2) in terms of being born in the Netherlands ($\chi 2$ (1, n=238) = 2.40, $p>.05$) and gender ($\chi 2$ (3, n=238) = 1.91, $p>.05$). There was a significant difference between the conditions in terms of how many students spoke other languages than Dutch at home, with group 2 showing significantly fewer students who spoke an additional language at home ($\chi 2$ (1, n=238) = 6.67, $p = 0.01$), compared to group 1. For the other variables related to cultural background, too many assumptions for Chi squared tests were violated to provide a useful statistic. See Table 2 for an overview.

Table 2. Overview of descriptive statistics for participant characteristics over both conditions (group 1 and 2).

	Group 1 (n=159)	Group 2 (n=79)
Gender male	76 (47,8%)	35 (44,3%)
Gender female	77 (48,4%)	43 (54,4%)
Gender other / unknown	6 (3,8%)	1 (1,3%)
Born in Netherlands	144 (90,6%)	76 (96,2%)
Speaking other languages at home	33 (20,8%)	6 (7,6%)
Mother born in Netherlands	131 (82,4%)	71 (89,9%)
Father born in Netherlands	141 (88,7%)	72 (91,1%)

2.4 Missing data

Because not all of the students were able to participate in all measurement moments, attrition rates vary over the different times. Attrition was particularly high at M3 in the control condition. This was partly due to one teacher who forgot to administer the final task. See Appendix B for more details.

2.5 Pretest and posttests

2.5.1 Test versions

In each of the conditions students took part in three tests via Qualtrics in accordance with the switching replications design: a pretest, and two posttests. At each moment of measurement, students were digitally presented with one of three versions of a writing task randomly, consisting of two short emails in which they had to pose a request to another person. All the necessary contextual information to write the email was provided. The versions were counterbalanced over the three different measurement moments to avoid score differences that could solely be attributed to the writing task. See Appendix A for the writing tasks. All tasks were constructed in collaboration with a teacher and pretested to ensure that they were fitting for the target group.

In each version, the power distance and social distance between the writer and the addressee were kept constant. The difference between the tasks was in the rate of imposition: in one of the emails ('low R'), the rate of imposition was low (e.g., asking your team leader to switch a shift for a supermarket side job that would not have negative consequences for the company), in the other (' high R'), the rate of imposition was high (e.g., asking your team leader if you can switch shifts in your supermarket side job even though you know that you were specifically needed on the day that you want to switch out of). Students did not only randomly start with a different version, they also randomly had to write the high or the low rate of imposition mail first. They were instructed to write no more than 150 words per email. After each email, students were asked to indicate how difficult they felt writing the email was (self-efficacy), as self-efficacy beliefs correlate strongly with writing competence (Bruning et al., 2013). They were also asked to indicate how much mental effort the task had taken them, using the validated Mental Effort Rating Scale (MERS), a one-item scale ranging from 1-9 on which 1 indicated really little effort and 9 an enormous amount of effort (Paas, 1992). We hypothesized that mental effort would decrease as a result of the intervention, and that self-efficacy would increase.

2.5.2 Assessing students' writing

Each of the students' written requests (6 in total, from three different moments in time) were evaluated by a panel of text raters using *Comproved* software (https://www.comproved.com), which is based on comparative judgment (CJ): a highly-reliable method for assessment in which texts are assessed through pair-wise comparison. Recent studies have shown that comparative judgement is more valid and reliable than other forms of assessment, such as using rubrics (Pollitt, 2012; Sadler, 2009; Van Daal et al., 2022; Verhavert et al., 2019). The method is also considered intuitive and easier to use as raters base their decisions on comparisons rather than analyzing each product in an absolute manner (Laming, 2003); assessing each

text is also estimated to take up to five times less time than in other forms of assessments (Coertjens et al., 2017), which is a great advantage when dealing with large amounts of texts such as in the current study. Additionally, CJ is not susceptible to sequential effects and reduces specific raters' biases (Pollitt, 2012).

All 1263 texts that we obtained were entered into *Comproved* manually, where raters were asked to answer one overarching question when comparing two randomly clustered texts: which of these texts is most appropriate in terms of politeness, given the context? The reason for phrasing the question this was is that simply asking the question 'which text is most polite?' would be potentially problematic, as students could also use too much politeness for the context. Rather than aiming to capture the most polite text, the assessment thus intended to capture the most *pragmatically appropriate* use of politeness.

Internal consistency in CJ assessments is most commonly determined by the Scale Separation Reliability (SSR), reflecting the consistency between raters in general. The outcome is thus a text ranking in which each text is placed, ranging from roughly -5 to 5, with 0 as the mean text score (see Figure 2). All texts with values below 0 are thus considered below average; all texts with scores above 0 are considered above. In the current assessment, the texts were rated by a panel of 50 raters, who were aware of the goals of the study, but who had no idea which text belonged to which condition, school or moment of measurement, nor did they receive specific instructions about what to look out for in the assessment. Of these raters, 22 were master's students in a related field (communication, linguistics or teacher education), 15 were teacher educators with relevant expertise, 8 were Dutch language teachers (who were not involved in the project otherwise), 2 were full-time researchers in language and education and 3 had relevant academic backgrounds but could not be classified in any of the aforementioned groups. Rating one pair of texts took about 22 seconds (median). Each rater rated anywhere between 3 and 600 texts (*Mean* = 153,9; *Median* = 95; *SD* = 162,4), depending on their available time. In the current study, each text was compared to another text 12.2 times on average, resulting in a sufficiently reliable SSR of .71 (Verhavert et al., 2019).

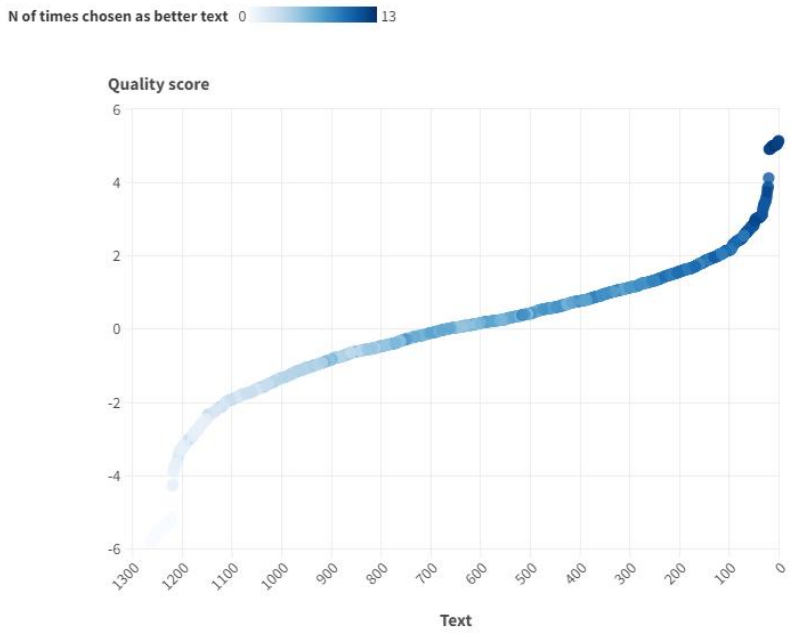
2.6 Implementation fidelity

To gain some insights into the question whether teachers implemented the intervention as intended (see O' Donnell, 2008), we asked them to report whether each lesson had gone according to plan in a fidelity form, outlining any major deviations to us. While all teachers had performed the intervention mostly as intended, none of the teachers managed to perform all of the tasks in class, due to various circumstances: mostly changes in the teachers schedule or because the intervention was a bit much.

We also checked intervention enactment by examining how much of the tasks from the intervention booklet were completed by the students. On average, 81,3% of the assignments were completed (SD = 16,8), based on 162 booklets that were

returned to us (68%). This percentage of completion is just above the 80% minimum that Muijs and Reynolds (2010) consider crucial for effective teaching.

Figure 2. Text ranking based on CJ with quality scores on the Y-axis. 0 is the mean score. The darker the red, the higher the number of times the text was chosen as the better one in the pair-wise comparison.



2.7 Data analysis

Due to the nested nature of the data, with students nested within classrooms and writing tasks nested within students, we employed multilevel modeling for our analysis (Sommet & Morselli, 2007). This analysis was conducted using the GAMLj package (Gallucci, 2019) within the Jamovi software (The Jamovi Project, 2022). The intraclass correlation (ICC) indicated that *classroom* could only account for 3,54% of the variance of writing scores; *students* accounted for 38,7% of the variance. For each of the six dependent variables *score* (high and low R score), *self-efficacy* (self-efficacy high and low R) and *mental effort* (mental effort high and low R), we adopted a step-wise build-up strategy (Glas et al., 2007), beginning with a basic intercept-only model (Model 1) and using Maximum Likelihood estimation. In this model, we estimated the intercept (i.e., mean score) of each variable, as well as estimation of variance of class, student and the residuals (i.e., the within-person variation over time). In Model 2, a fixed effect of Time was added, to see whether changes over time took place regardless of condition. Model 3 then also took Condition into account, where

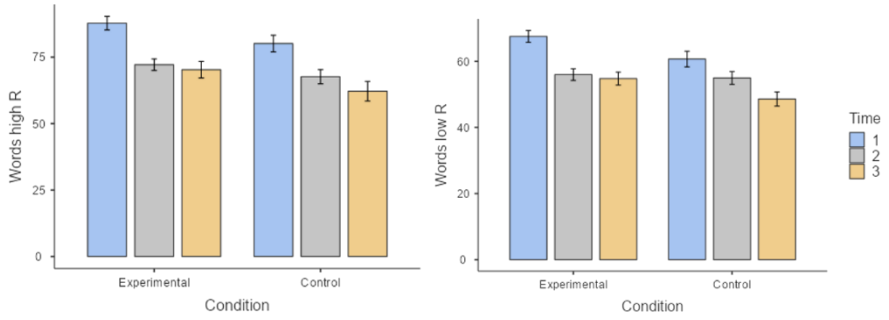
the interaction between Time*Condition was of primary interest. Finally, given the decrease in the number of words over time, Model 4 took into account the number of words for each writing task. The best model for each dependent variable was determined by comparing the -2 log likelihood values of the extended model compared to the previous model by means of a χ2 test.

3. RESULTS

3.1 Descriptive statistics

Students wrote an average of 75,5 words for the high R texts (*SD* = 22,2; *range*: 16,3-160). For the low R texts, the mean number of words was 58,5 (*SD* = 18,1; range 14,3-144), which was theoretically expected as lower rates of imposition require less politeness compared to higher rates. Bar plots suggest that the mean number of words declined over time (see Figure 3).

Figure 3. Mean n of words for high R and low R tasks per condition per measurement moment (time). Error bars represent standard errors.



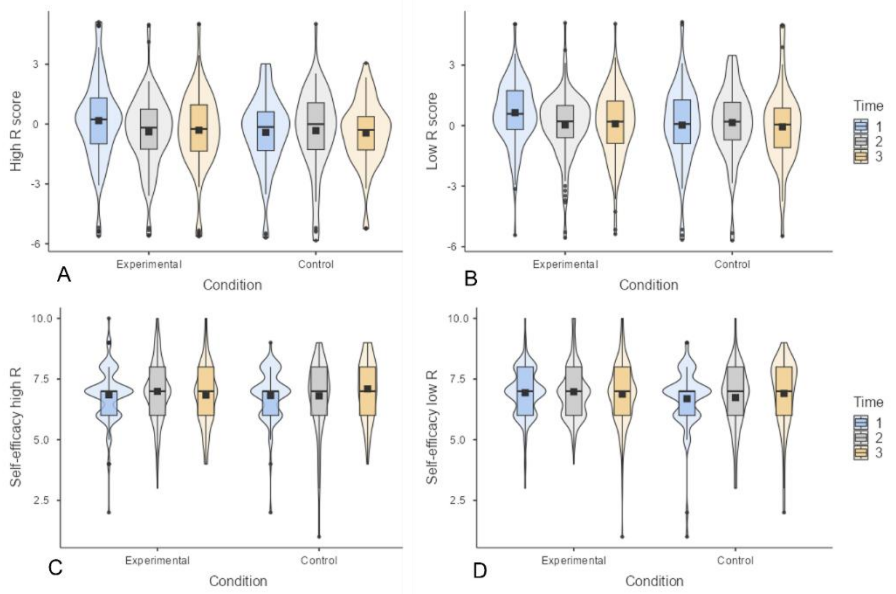
The violin plots in Figure 4 show students' writing scores and self-efficacy for high and low rates of imposition (high/low R); those in Figure 5 show students' perceived mental effort (MERS). From these data, the impression arises that scores for all of these variables seem to remain fairly static over time for both conditions. See Appendix B for more details.

3.2 Multilevel analyses

Table 3 presents model comparisons for the multilevel repeated measures design related to high R tasks, while Table 4 focuses on low R tasks. Tables 5 and 6 provide the parameter estimates for the final models, and Figure 6 visualizes the scores of these models for all dependent variables. The model information indicates that for all dependent variables, except low R scores, there is no significant interaction

between Time and Condition (M3), as the models do not improve with the addition of this variable. For low R scores, however, a significant interaction was observed (p = 0.022). Despite this significant interaction effect in the multilevel model for low R scores (Time1*Condition1, M3) shown in Table 4, post hoc comparisons revealed no significant pairwise differences between time points and conditions after applying the Bonferroni correction for multiple comparisons. This discrepancy likely arises because the interaction effect in the multilevel model reflects an overall pattern across groups, whereas the post hoc tests assess specific pairwise differences.

Figure 4. Violin plots showing mean scores (black square), distribution and density of the data for the writing scores (panel A: high R, panel B: low R) and for self-efficacy scores (panel C: high R, panel D: low R) for both conditions.



For most cases, except for the mental effort variable in high R tasks, adding the number of words significantly improved the models (M4), indicating that longer texts tend to be evaluated more favorably. Pearson's correlations support this, showing a strong correlation between quality scores and the number of words for high R tasks (r(615) = 0.54, p< .001) and a moderate correlation for low R tasks (r(617) = 0.43, p< .001). The number of words also appears to influence self-efficacy and mental effort scores. However, controlling for the number of words does not result in a significant Time*Condition interaction. For both high and low R tasks, subsequent multilevel models show that Time is a significant predictor for the number of words (p<.001): according to a post-hoc test, the mean number of words for low R-tasks declines significantly between moment 1 and 2 (p <.001), between moment 1 and 3 (p <.001) and between 2 and 3 (p = 0.038). For high R tasks, the same effect is found, with one

difference: while there is a significant decline between moment 1 and 2 (p <.001) and 1 and 3 (p <.001), there is no significant decline between moment 2 and 3 (p = 0.48). For the mean number of words, no significant differences were found between conditions, nor were there interaction effects between Time and Condition. This decline in the number of words is visualized in Figure 7, and they confirm the suspicion that we had based on Figure 3.

Figure 5. Violin plots showing mean scores (black square), distribution and density of the data for the perceived mental effort (MERS) (panel E for high R texts, panel F for low R texts)

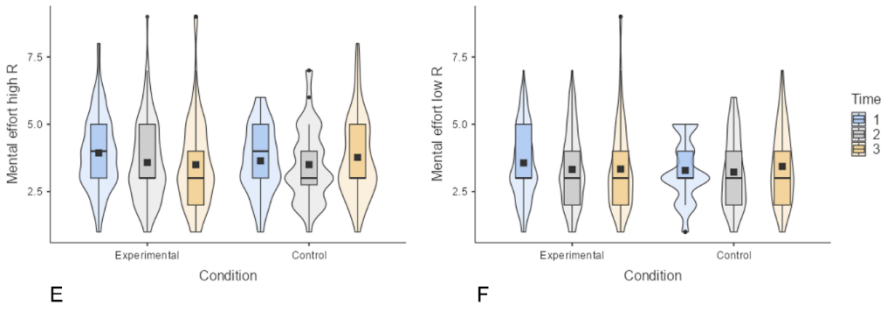


Figure 6. Scores on the different dependent variables, with the A-panel showing scores for high rate of imposition texts, and the B panel showing scores for low ones.

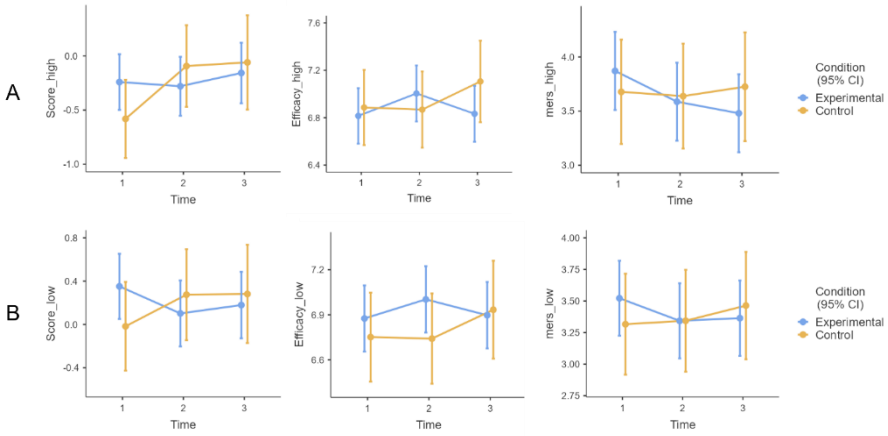


Table 3. comparisons for the multilevel repeated measures design for the variables related to high R tasks

	High R mental effort					High R self-efficacy					High R text score				
	-2LL	Δ-2LL	df	p	AIC	-2LL	Δ-2LL	df	p	AIC	-2LL	Δ-2LL	df	p	AIC
M1	-1267.6	-	-	-	2543.3	-943.1	-	-	-	1894.3	-1044.3	-	-	-	2096.7
M2	-1264.3	3.3	2	0.03	2540.5	-942.8	0.3	2	0.75	1897.7	**-1038.6**	**5.7**	**2**	**0.003**	**2089.3**
M3	-1261.9	2.4	5	0.17	2541.9	-940.0	2.8	5	0.06	1897.8	-1036.7	1.9	5	0.147	2091.4
M4	-1164.0	**97.9**	**6**	**<0.01**	**2347.9**	**-907.9**	**32.1**	**6**	**0.05**	**1835.9**	-1018.7	18.0	6	0.119	2057.4

Table 4. Comparisons for the multilevel repeated measures design for the variables related to low R tasks

	Low R mental effort					Low R self-efficacy					Low R text score				
	-2LL	Δ-2LL	df	p	AIC	-2LL	Δ-2LL	df	p	AIC	-2LL	Δ-2LL	df	p	AIC
M1	-1180.9	-	-	-	2369.7	-943.4	-	-	-	1894.9	-1015.8	-	-	-	2039.6
M2	-1174.8	6.1	2	0.002	2361.7	-943.3	0.1	2	0.90	1898.7	-1014.0	1.8	2	0.17	2040.0
M3	-1170.7	4.1	5	0.022	2359.4	-941.3	2.0	5	0.32	1900.5	-1012.9	1.1	5	0.36	2043.7
M4	**-1118.0**	**52.7**	**6**	**<0.01**	**2256.1**	**-913.8**	**27.5**	6	**0.002**	**1847.5**	**-990.3**	**22.6**	**6**	**0.01**	**2000.5**

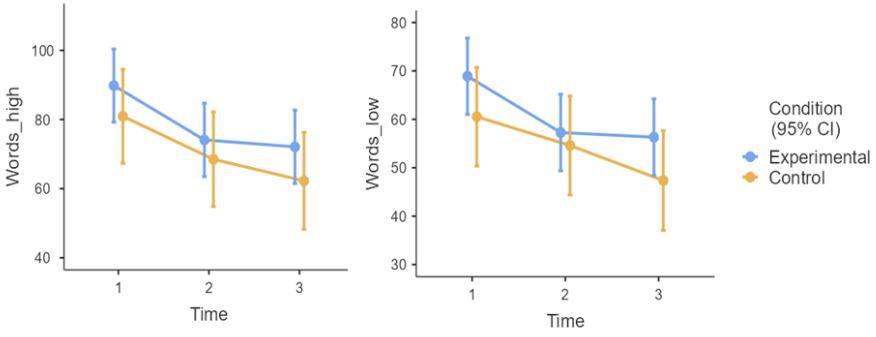
Table 5. Parameter estimates (Est.), Standard errors (St. err.) and p-values based on Model 4 for high R dependent variables

	High R mental effort			High R self-efficacy			High R text score		
	Est.	St. err.	p	Est.	St. err.	p	Est.	St. err.	p
Intercept	-0.235	0.158	0.002	6.920	0.075	< .001	3.663	0.119	< .001
Time1 (2-1)	0.225	0.169	0.156	0.086	0.089	0.336	-0.162	0.103	0.117
Time2 (3-1)	0.303	0.152	0.074	0.119	0.097	0.224	-0.172	0.113	0.129
Condition (C-E)	-0.019	0.002	0.902	0.069	0.151	0.664	0.034	0.238	0.890
Words	0.033	0.309	< .001	0.004	0.001	0.005	0.003	0.002	0.119
Time*Condition 2-1 (C-E)	0.527	0.329	0.089	-0.207	0.173	0.233	0.200	0.200	0.222
Time*Condition 3-1 (C-E)	0.438	0.158	0.183	0.202	0.188	0.283	0.438	0.217	0.044

Table 6. Parameter estimates (Est.), Standard errors (St. err.) and p-values based on Model 4 for low R dependent variables

	Low R mental effort			Low R self-efficacy			Low R text score		
	Est.	St. err.	p	Est.	St. err.	p	Est.	St. err.	p
Intercept	0.195	0.095	0.086	6.867	0.067	<0.01	3.392	0.096	<.001
Time1 (2-1)	0.021	0.133	0.873	0.059	0.092	0.523	-0.076	0.102	0.457
Time2 (3-1)	0.063	0.144	0.662	0.102	0.101	0.313	-0.005	0.111	0.963
Condition (C-E)	-0.031	0.191	0.876	-0.116	0.133	0.430	-0.035	0.192	0.861
Words	0.033	0.003	<0.001	0.007	0.002	0.002	0.004	0.002	0.010
Time*Condition 2-1 (C-E)	0.543	0.262	0.039	-0.137	0.179	0.444	0.206	0.198	0.297
Time*Condition 3-1 (C-E)	0.472	0.277	0.089	0.160	0.193	0.406	0.304	0.214	0.154

Figure 7. Visualization of the decline in the number of words written for both conditions over time, for both high (left panel) and low R tasks (right panel).



4. CONCLUSION AND DISCUSSION

The current study set out to explore how concepts from pragmatics could be introduced meaningfully into L1 language education, and whether this may influence students' writing. To the best of our knowledge, this is the first study of its kind in the realm of L1 language education. To re-iterate, we attempted to answer the following research question: To what extent does a short intervention targeting explicit pragmatics concepts (speech acts, conversational implicature and politeness) in an L1 context benefit pre-university students' written requests? Based on the analyses that we've presented here, it might be tempting to conclude that the intervention in its current form does not benefit students' writing of requests. Indeed, the data show that none of the dependent variables show any signs of improvement as a consequence of the intervention: scores do not go up, self-efficacy does not improve and perceived mental effort does not decline. Given that the majority of students underwent the intervention roughly as intended (based on our fidelity measures), it seems that the current treatment was insufficient to cause any serious changes to come about. It also seems unlikely that our sample was not big enough to find effects in the first place, as our study was adequately powered.

One crucial factor that likely influenced these results is the significant decline in the number of words written over time. Pragmatics theory suggests that tasks with a high rate of imposition should theoretically require more words, not fewer. Therefore, this observed decrease in word count is unexpected and warrants closer examination. We interpret this reduction in the mean number of words as indicative of decreasing motivation for the writing tasks, a common issue in low-stakes testing (Cohen et al., 2011, p. 211). Students may have perceived the tasks as either too easy or too repetitive over time, or they may not have taken the tasks seriously due to the absence of grading. Despite the relationship between text length and text quality, this lack of motivation might explain why the scores remain unaffected.

In an attempt to mitigate this problem, we designed the tasks to involve writing two different (but similar) requests at each measurement point. However, it is possible that these tasks were not sufficiently distinct to maintain student interest. Future research should explore the development of more varied tasks and emphasize the importance of teachers clearly explaining the goal of the writing tasks to enhance student motivation (Cohen et al., 2011). Another potential solution to address declining motivation is to reconsider the switching replications design, which required students to write six texts (two per measurement), as opposed to a classical experimental-control condition that uses only two measurements (and thus four texts). However, compromising on the design may not be advisable since the switching replications design offers greater methodological rigor than a simple pre-test-posttest design (see introduction). Furthermore, other studies have shown that a decline in motivation for writing tasks does not necessarily occur within such designs (e.g., Bouwer et al., 2018; Elving, 2019).

In spite of the potential impact of test motivation on the results, it might also be the case that the intervention itself was too short to facilitate a proper transfer from theoretical concepts to actual writing, although even short treatments have proven to be fruitful in L2 contexts (Ren et al., 2023). Fully appreciating abstract disciplinary concepts such as *face keeping* might require more learning time, or, if transfer is the goal, a different type of core tasks, based on more refined design principles. The current design principles were based on other short interventions in the linguistic domain. Pragmatics does not necessarily thrive under the same types of design principles, so a deeper exploration of effective design principles in a qualitative manner would be useful for future studies. Future studies may also re-visit our decision to incorporate several pragmatic concepts into one intervention. While the intervention was centered around pragmatic politeness, two of the four lessons revolved around notions building up to politeness, such as *speech acts* and *implicature*, which may have taken away too much time from politeness theory itself. The intervention may have introduced too many concepts to students, thus limiting their capacity to really absorb all of these notions.

Similarly, we have not tested to what extent students have obtained a real understanding or knowledge of the pragmatic concepts that we introduced. Follow-up studies would do well to take this into account, so that we can better assess whether transfer occurs. In doing so, researchers should be aware that this will increase the test load at each moment of measurement, so efforts would have to be taken to minimize testing fatigue (Borgonovi & Biecke, 2016). In our study, some key concepts—novel to all participants—may not have been fully grasped. This likely applied to both students and teachers, given their limited exposure to disciplinary linguistics (see our description of the Dutch context). Moreover, lay interpretations of politeness may have interfered with a more technical, linguistic understanding (see Brown, 2015). It is plausible that students assumed familiarity with the concept due to their everyday encounters with politeness, leading them to perceive the intervention as reiterating known ideas. Since politeness features prominently in daily life and the

standard curriculum, students may have underestimated the novelty of the pragmatic perspective presented. While the intervention distinguished between everyday politeness and its pragmatic counterpart, additional time may be necessary to facilitate deeper engagement with politeness from a disciplinary standpoint and to internalize related pragmatic concepts.

While most students have completed most of the tasks from their booklet, the vast majority have skipped a handful of core assignments that required a lot of writing, such as tasks in which they had to summarize what they had learned about certain concepts. For example, in one task, students were asked to explain what they had learned that lesson about speech acts, and they were asked to use the following concepts in their answer: *performative verb, locution, illocution, perlocution*. Most of the students did not complete this task, simply leaving it open. Likewise, tasks in which they were encouraged to *reason* about an answer (design principle 4) were oftentimes completed, but with only a limited amount of reasoning being visible in writing. This aligns with students' lack of motivation in the test tasks, where their willingness to write larger bodies of texts also appeared to be low. This illustrates that we need a better understanding of how such interventions are delivered in practice beyond simply relying on teachers' fidelity reports and keeping track of the number of tasks students have performed. In this sense, while we did adhere certain aspects of fidelity (focusing on adherence, participant responsiveness and duration by means of fidelity logs), we did not consider fidelity measures related to quality of delivery and program differentiation (O' Donnell, 2008). In follow-up studies, improved fidelity measures are thus essential. Data on how students engage with learning during the intervention are essential for understanding why certain tasks may be skipped and what conceptual barriers arise when reasoning about pragmatic concepts. Equally important is examining how teachers deliver the intervention, as this likely shapes students' learning processes and challenges.

The teachers did report that most of the students seem to like the intervention material, and the teachers themselves had fun working with it. This means that the intervention does have educational potential, even though we did not find any statistical effects of the intervention on the target variables. Future projects should thus explore how this 'fun' factor can be retained (as this probably impacts motivation in a positive way), whilst also looking for ways to stimulate students to write more, both in the intervention and in the tests.

The current study has been unable to conclusively determine whether a short introduction of explicit pragmatic concepts might benefit students' request writing, given the decline in the number of words written. Nevertheless, it is important for these results to be shared, if only to avoid a publication bias in which only positive results are reported in academic journals (e.g., Ritchie, 2020, ch. 4), which Ren and colleagues (2023) also found some evidence for in their meta-analysis. Negative results are published considerably less often than positive ones because reviewers are less likely to recommend publication for such submissions (Hopewell et al., 2009). The study also offers valuable recommendations for future studies in examining the

added value of introducing pragmatic concepts to L1 language education, especially in light of curriculum reforms in which more emphasis is given to knowledge *about* language or communication, such as the Dutch curriculum.

FUNDING

This work was supported by a Veni-grant awarded to dr. J.H.M. van Rijt by the Dutch Research Council (NWO) under Grant VI.Veni.221C.018.

REFERENCES

Austin, J. (1962). *How to do things with words*. Oxford University Press.

Bax, S., Kroon, S., & Spotti, M. (2024). Afterword - L1 Education between von Humboldt and Chat GPT. *L1-Educational Studies in Language and Literature, 24*(2). https://doi.org/10.21248/l1esll.2024.24.2.749

Bernstein, B. (2003). *Structuring of Pedagogic Discourse*. Taylor & Francis.

Biesenbach-Lucas, S. (2007). Students writing emails to faculty: an examination of e-politeness among native and non-native speakers of English. *Language Learning and Technology, 11*(2), 59-81.

Borgonovi, F. & Biecek, P. (2016). An international compariosn of students' ability to endure fatigue and maintain motivation during a low-stakes test. *Learning and Individual Differences, 49*, 128-137. https://doi.org/10.1016/j.lindif.2016.06.001

Bouwer, R., Koster, M., & van den Bergh, H. (2018). Effects of a Strategy-Focused Instructional Program on the Writing Quality of Upper Elementary Students in the Netherlands. *Journal of Educational Psychology 110*(1), 58-71. https://doi.org/10.1037/edu0000206

Brown, P. (2015). Politeness and language. In: Wright, J. (Ed.) *International Encyclopedia of the Social and Behavorial Sciences (Second edition)* (pp.326-330). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.53072-4

Brown, P. & Levinson, S. (1987). *Politeness: some universals in Language Use*. Cambridge University Press.

Bruning, R., Dempsey, M., Kauffman, F., McKim, C. & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology, 105*(1), 25-38. https://doi.org/10.1037/a0029692

Coertjens, L., Lesterhuis, M., Verhavert, S., Van Gasse, R., & De Maeyer, S. (2017). Teksten beoordelen met criterialijsten of via paarsgewijze vergelijking: een afweging van betrouwbaarheid en tijdsinvestering. *Pedagogische Studiën, 94*(4), 283–303.

Clark, B. (2016). Pragmatics. In: Giovanelli, M. & Clayton, D. (Eds.) *Knowing About Language: Linguistics and the Secondary English Classroom* (pp. 64-76).

Cohen, L., Manion, L. & Morrison, K. (2011). *Research methods in education.* Routledge.

Culpeper, J. 2011. *Impoliteness: Using Language to Cause Offence*. Cambridge University Press.

Cummings, L. (2005). *Pragmatics: A Multidisciplinary Perspective.* Edinburgh University Press.

Cutting, J. (2015). *Pragmatics. A resource book for students.* Routledge.

Maeyer (2021, March 24). *Reproducible Stats in Education Sciences: Time to switch?*. Retrieved from https://svendemaeyer.netlify.app/posts/2021-03-24_Time-to-Switch/

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). *Linguistics at School. Language Awareness in Primary and Secondary education*. Cambridge University Press.

Dera, J., Gubbels, J., Loo, J. V. D., & Van Rijt, J. (2023). *Vaardig met Vakinhoud: Handboek Vakdidactiek Nederlands*. Coutinho.

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*(1/2), 182-206. https://doi.org/10.1075/dujal.19038.die

Dols-Koot, I., & Van Rijt, J. (2019). 'Hakken en plakken': Een verkenning van de rol van morfologisch bewustzijn in het spellingonderwijs. *Dutch Journal of Applied Linguistics, 8*(2), 310-330. https://doi.org/10.1075/dujal.17030.dol

Economidou-Kogetsidis, M. (2011). ''Please answer me as soon as possible'': Pragmatic failure in non-native speakers' e-mail requests to faculty. *Journal of Pragmatics, 43*, 3193-3215. https://doi.org/10.1016/j.pragma.2011.06.006

Economidou-Kogetsidis, M. (2016). Variation in evaluations of the (im)politeness of emails from L2 learners and perceptions of the personality of their senders. *Journal of Pragmatics, 106 43*, 1-19. https://doi.org/10.1016/j.pragma.2016.10.001

Elving, K. (2019). *Effectieve leeractiviteiten voor het schrijfonderwijs in 4 havo*. Doctoral thesis. Amsterdam: LOT.

Gallucci, M. (2019). *GAMLj: General analyses for linear models*. [jamovi module],' available at https://gamlj.github.io/. Accessed 15 June 2024.

Goffman, E. (1967) On Face-Work. In: Goffman, E., (Ed.), *Interaction Ritual, Pantheon*, New York, 5-45.

Green, B. & Erixon, P.-O. (Eds.) (2020). *Rethinking L1 Education in a Global Era. Understanding the (Post-)National L1 Subjects in New and Difficult Times*. Springer.

Grice, P. (1975). Logic and conversation. In: P. Cole & J. Morgan (Eds.) *Pragmatics (syntax and semantics)* 9. Academic Press.

Hopewell, S., Loudon, K., Clarke, M. J., Oxman, A. D., & Dickersin, K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews*, 2009(1), MR000006. https://doi.org/10.1002/14651858.MR000006.pub3

Hordern, J. (2021). Recontextualisation and the teaching of subjects. *The Curriculum Journal, 32*(4), 592-606. https://doi.org/10.1002/curj.110

Horn, L. (2007). Neo-Gricean pragmatics: a Manichaean manifesto. In: N. Burton-Roberts (Ed.) *Pragmatics* (pp. 158-183). Pallgrave Macmillian.

Houtkoop, H. & Koole, T. (2008). *Taal in actie. Hoe mensen communiceren met taal*. Coutinho.

Hudson, R. (2004). Why education needs linguistics (and vice versa). *Journal of Linguistics, 40*(1), 105-130. https://doi.org/10.1017/S0022226703002342

Laming, D. (2003). *Human judgment: The eye of the beholder* (1st ed.). Cengage Learning EMEA.

Levinson, S. (2000). *Presumptive meanings*. Cambridge University Press.

Moseley, D., Elliott, J., Gregson, M. & Higgins, S. (2005). Thinking skills frameworks for use in education and training. *British Educational Research Journal*, 31(3), 367-390. https://doi.org/10.1080/01411920500082219

McConachy, T. (2013). Exploring the meta-pragmatic realm in English language teaching. *Language Awareness, 22*(2), 100-110. https://doi.org/10.1080/09658416.2011.644795

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), 139-166. https://doi.org/10.1080/02671522.2011.637640

Muijs, D., & Reynolds, D. (2010). *Effective teaching: Research and practice*. Paul Chapman.

O'Donnell, C. L. (2008). Defining, Conceptualizing, and Measuring Fidelity of Implementation and Its Relationship to Outcomes in K–12 Curriculum Intervention Research. *Review of Educational Research, 78*(1), 33-84. https://doi.org/10.3102/0034654307313793

Paas, F. (1992).Training strategies for attaining transfer of problem-solving skills in statistics: a cognitive-load approach. *Journal of Educational Psychology, 84*(4): 429-434. https://doi.org/10.1037//0022-0663.84.4.429

Pollitt, A. (2012). Comparative judgement for assessment. *International Journal of Technology and Design Education, 22*(2), 157–170. https://doi.org/10.1007/s10798-011-9189-x

Ren, W., Li, S. & Lu, X. (2023). A meta-analysis of the effectiveness of second language pragmatics instruction. *Applied Linguistics, 44*(6), 1010–1029. https://doi.org/10.1093/applin/amac055

Ritchie, S. (2020). *Science fictions. Exposing Fraud, Bias, Negligence and Hype in Science*. The Bodley Head.

Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 1–19). Dordrecht: Springer.

Sawyer, W., & Van den Ven, P.-H. (2007). Starting points: Paradigms in mother tongue education. *L1-Educational Studies in Language and Literature, 7*(1), 5–20. https://doi.org/10.17239/L1ESLL-2007.07.01.06

Searle, J. (1969). *Speech acts.* Cambridge University Press.

Shadish, W., Cook, T. and Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company.

Sheehan, M., Corr, A., Havinga, A., Kasstan, J. & Schifano, N. (2021). Rethinking the UK languages curriculum: Arguments for the inclusion of linguistics. *Modern Languages Open*, 1–24. https://doi.org/10.3828/mlo.v0i0.368

Sinha, T. & Kapur, M. (2021). When problem solving followed by instruction works: evidence for productive failure. *Review of Educational Research, 91*(5), 761-798. https://doi.org/10.3102/00346543211019105

SLO (2024). *Definitieve concept-kerndoelen Nederlands*. SLO. https://www.slo.nl/@24174/definitieve-conceptkerndoelen-nederlands/

Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology, 30*(1), 203–218. https://doi.org/10.5334/irsp.90

Song, S. (2017). The Brown and Levinson theory revisited: A statistical analysis. *Language Sciences, 62*, 66-75. https://doi.org/10.1016/j.langsci.2017.03.006

Steenbakkers, J. (2023). Speelse toewijding. Een pedagogisch-didactisch onderzoek naar schrijfstijl en formuleeronderwijs in klas 3 en 4 havo/vwo. Doctoral dissertation. Rijksuniversiteit Groningen.

Taguchi, N., Chen, Y. & Qin, Y. (2023). Metapragmatic knowledge and transfer of learning across speech acts. *Applied Linguistics, 44*(2), 217-238. https://doi.org/10.1093/applin/amac043

Thomas, J. (1995). *Meaning in interaction*. London: Longman.

Van Daal, T., Lesterhuis, M., De Maeyer, S., & Bouwer, R. 2022). Validity,reliability and efficiency of comparative judgement to assess student work. *Frontiers in Education, 7*:1100095. https://doi.org/10.3389/feduc.2022.1100095

Verhavert, S., Bouwer, R., Donche, V., & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy, & Practice, 26*(5), 541-562. https://doi.org/10.1080/0969594X.2019.1602027

Van der Wijst, P. (1996). *Politeness in requests and negotiations*. Doctoral dissertation. ICG.

Van Rijt, J. (2022). Taalkunde in de bovenbouw van havo en vwo: De vormgeving van een betwist vakonderdeel [Linguistics in the upper years of havo and vwo: The design of a contested subject component]. *Levende Talen Tijdschrift, 23*(1), 36–47.

Van Rijt, J. (2024). Learning how to think like a linguist: Linguistic reasoning as a focal point in L1 grammar education. *Language and Linguistics Compass, 18*(3), 1-17. https://doi.org/10.1111/lnc3.12513

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., & Coppen, P.-A. (2021). The conceptual importance of grammar: Knowledge-related rationales for grammar teaching. *Pedagogical Linguistics, 2*(2), 175-199. https://doi.org/10.1075/pl.21008.van

Van Rijt, J., Myhill, D., De Maeyer, S., & Coppen, P.-A. (2022). Linguistic metaconcepts can improve grammatical understanding in L1 education evidence from a Dutch quasi-experimental study. *PLoS One, 17*(2), Article e0263123. https://doi.org/10.1371/journal.pone.0263123

Van Rijt, J., Wijnands, A., & Coppen, P.-A. (2020). 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/10.1080/09500782.2019.1690503

Wegman, H., & van Rijt, J. (2023). Taal in gebruik: Pragmatiek in de klas. In J. Dera, J. Gubbels, J. van der Loo, & J. van Rijt (Eds.), *Vaardig met vakinhoud: Handboek vakdidactiek Nederlands* (pp. 209-220). Coutinho.

Young, M. (2013). Overcoming the crisis in curriculum theory: a knowledge-based approach. *Journal of Curriculum Studies, 45*(2), 101-118. https://doi.org/10.1080/00220272.2013.764505

APPENDIX A. WRITING TASK (TRANSLATED FROM DUTCH)

	High rate of imposition	Low rate of imposition
1	**Write an email to the principal of your school.** You are consciously thinking about the future of the planet, and have therefore decided to eat only plant-based foods. You have noticed that there are very few vegan products available in the school cafeteria. You would like the selection of products in the cafeteria to change, and for more plant-based options to be made available. You are writing an email to the principal of the school to make this happen. The name of the principal is Carla Frederiks.	**Write an email to the principal of your school.** You participate in gymnastics at a top sport level. Next month, there is an important competition that you want to participate in. The competition is on a Friday, so you will need time off from school. At your school, there is a policy that students who do top-level sports can generally get time off. However, you always still need to ask for permission. Normally, you would ask your mentor, but they are not available, so you are writing to the principal. Her name is Carla Frederiks.
2	**Write an email to the treasurer of your hockey club. (**The treasurer is the person responsible for the finances of the association, i.e., the financial matters.) You are a member of a hockey club, and you are talented. Because your parents cannot afford the membership fee (contribution), you pay it yourself. To earn this money, you work at a supermarket. Due to your talent, you have been invited to train with the adults starting next season. However, this would mean you have to pay a higher membership fee, specifically the adult rate (80 euros per month). This amount is actually too high for you, and you would like to continue paying the same amount as you do now (40 euros per month). Therefore, you are writing an email to the club's treasurer. The treasurer of your hockey club is named Florien Martens**.**	**Write an email to the treasurer of your hockey club. (**The treasurer is the person responsible for the finances of the association, i.e., the financial matters.) You are a member of a hockey club and pay 40 euros every month for your membership (contribution). Last month, this amount was debited from your bank account twice, even though you only need to pay once per month. You would like the money to be refunded, so you are writing an email to the treasurer of the club. The treasurer of your hockey club is named Florien Martens.
3	**Write an email to the team leader of your part-time job at the supermarket.** You work in a supermarket every Saturday. When you were hired, the team leader was specifically looking for someone who could work on Saturdays, as there were not enough employees available on that day. However, it is no longer convenient for you to work on Saturdays because you now have hockey matches on that day. Instead, you would prefer to work on Sundays. You are writing an email to your team leader to request this change. Your team leader's name is Roy Nieuwenhuis.	**Write an email to the team leader of your part-time job at the supermarket.** You work in a supermarket every Saturday. Next week, you don't want to work on Saturday because it's your birthday. You've arranged with your colleague Myrthe to swap shifts: she will work for you on Saturday, and you will take her place on Sunday. You just need approval from your team leader, so you are writing him a message. Your team leader's name is Roy Nieuwenhuis.

APPENDIX B. DESCRIPTIVE STATISTICS FOR THE RELEVANT VARIABLES (INC. MISSING DATA)

	Condition	Time	N	Missing	Mean	Median	SD
High R score	Experimental	1	157	2	0.172	0.234	2.060
		2	134	25	-0.388	-0.176	1.796
		3	127	32	-0.306	-0.244	1.888
	Control	1	77	2	-0.414	-0.145	1.817
		2	70	9	-0.326	-0.003	2.046
		3	55	24	-0.445	-0.291	1.643
Self-efficacy high R	Experimental	1	156	3	6.853	7	1.064
		2	133	26	6.993	7	1.234
		3	131	28	6.840	7	1.239
	Control	1	78	1	6.821	7	1.054
		2	72	7	6.806	7	1.469
		3	52	27	7.096	7	1.209
Low R score	Experimental	1	156	3	0.641	0.584	1.526
		2	136	23	0.033	0.211	1.542
		3	131	28	0.088	0.199	1.662
	Control	1	76	3	0.028	0.080	1.959
		2	70	9	0.151	0.197	1.742
		3	54	25	-0.066	0.052	1.936
Self-efficacy low R	Experimental	1	156	3	6.936	7	0.975
		2	134	25	6.978	7	1.093
		3	131	28	6.878	7	1.283
	Control	1	77	2	6.688	7	1.228
		2	72	7	6.736	7	1.267
		3	53	26	6.906	7	1.377
Mental effort high R	Experimental	1	156	3	3.936	4	1.385
		2	133	26	3.579	3	1.458
		3	131	28	3.496	3	1.647
	Control	1	78	1	3.641	4	1.238
		2	72	7	3.500	3	1.374
		3	52	27	3.769	3	1.477
Mental effort low R	Experimental	1	156	3	3.564	3	1.316
		2	134	25	3.321	3	1.369
		3	131	28	3.336	3	1.572
	Control	1	77	2	3.286	3	1.145
		2	72	7	3.222	3	1.335
		3	53	26	3.434	3	1.366