

Co-creating tools to monitor first graders' progress in reading: A balancing act between perceived usefulness, flexibility and workload

1 Eve Francotte^{1*}, Stéphane Colognesi^{1†}, Liesje Coertjens^{1†}

2 ¹Psychological Sciences Research Institute, UCLouvain, Louvain-la-Neuve, Belgium

3 †: Equal contribution and last authorship

4 * Correspondence:

5 Eve Francotte (she/her)

6 eve.francotte@uclouvain.be

7 **Keywords:** progress monitoring¹, formative assessment², practice-embedded research³,
8 teaching practices⁴, reading⁵

9 **Number of words:** 9637, number of tables: 3, number of figures: 4 (all figures are presented in
10 supplementary materials)

11 Abstract

12 Educational inequalities – i.e., the achievement gaps between pupils from disadvantaged
13 backgrounds and their peers from advantaged backgrounds – are present in many OECD countries.
14 This is particularly problematic in reading, which is a predictor of future academic and social
15 success. To reduce this reading achievement gap, recent meta-analyses point towards progress
16 monitoring: regularly measuring pupils' mastery levels and differentiating instruction accordingly.
17 However, the research recommendations only slowly make their way to teaching habits, particularly
18 because teachers may consider progress monitoring difficult and cumbersome to implement. To
19 avoid such difficulties, partnerships between teachers and researchers have been recommended.
20 These allow teachers' complex realities to be taken into account and, consequently, tools to be
21 designed that are meaningful and feasible for practitioners. Using an iterative and participatory
22 process inspired by practice-embedded research, the present research set out to (1) co-construct tools
23 to monitor first-graders' progress in reading, and, (2) examine how these tools met teachers' needs.
24 Five teachers in the French-speaking part of Belgium co-constructed four tools during four focus
25 groups. The transcribed discussions were analyzed using an interactional framework containing three
26 areas of knowledge: shared, accepted, and disputed. The results indicated three shared needs:
27 perceived usefulness, flexibility of the tools, and a desire to limit the workload. In addition, teachers
28 accepted that, between them, needs varied regarding the goal for progress monitoring and the format
29 of the evaluation. They had lengthy discussions on balancing workload and perceived utility, leading
30 them to conclude that there were two groups of teachers. The first group questioned the added value
31 of the progress monitoring tools in relation to their habitual practice. The second group on the other
32 hand described the added value for the teacher, certainly when aiming to grasp the level and
33 difficulties of struggling pupils. This second group had fewer years of teaching experience and
34 described their classroom practice as less organized compared to the teachers from the first group.
35 Theoretical and practical implications of these findings are discussed below.

1 Introduction

Multiple countries, such as France, Sweden, and Belgium, demonstrate strong educational inequalities, defined as the achievement gaps between pupils from disadvantaged and advantaged backgrounds (UNICEF Office of Research, 2016; Bricteux and Quittre, 2021). As reading predicts academic, professional, and social success (Slavin *et al.*, 2011; Oslund, Elleman and Wallace, 2021), this deficit has important consequences for pupils from disadvantaged backgrounds. Yet, reducing educational inequality is far from easy: in the French-speaking part of Belgium for example, despite governmental policy to reduce this inequality, the PIRLS results from 2011 and 2016 indicate a widening of the achievement gap (Schillings *et al.*, 2017).

The causes of this achievement gap were first explored via genetic or hereditary explanations, yet scientific studies do not confirm this hypothesis (for a review, see Nisbett *et al.*, 2012). Currently, the most accepted explanation concerns children's social and, particularly, home environments (Magnuson and Shager, 2010). Parents with higher educational attainment and more financial resources tend to provide their children with more conducive learning environments (Gennetian, Castells and Morris, 2010) and have higher academic expectations (Davis-Kean, 2005; Slates *et al.*, 2012). More specifically regarding language learning, pupils from advantaged backgrounds were found to interact more with their parents and use a larger vocabulary (Hart and Risley, 2003; Davis-Kean, 2005). As pupils' oral language skills are predictive of their reading skills (Le Normand, Parisse and Cohen, 2008; Bianco *et al.*, 2012), this results in significant differences between pupils, even before they enter primary school (Magnuson and Shager, 2010). However, these differences are not deterministic. Indeed, on average, in the Organisation for Economic Co-operation and Development (OECD), 11% of pupils have a resilient profile: they belong to the group of the 25% most disadvantaged pupils yet they reach the top 25% achievement in reading (Bricteux and Quittre, 2021). Furthermore, intervention studies have shown that it is possible to increase the reading skills of pupils from disadvantaged backgrounds (Dietrichson *et al.*, 2017, 2021) and of struggling readers more specifically (Slavin *et al.*, 2011; Neitzel *et al.*, 2021). Dietrichson and colleagues (2017) conducted a meta-analysis of interventions for pupils from disadvantaged backgrounds. Among the teaching practices studied, progress monitoring appears promising (*Hedges' g* = +0.32, 95%CI = [0.18; 0.47]) as does tutoring (*Hedges' g* = +0.36, 95%IC = [0.26; 0.45]). However, the latter has a higher cost per pupil in time than teaching practices aimed at the whole class (Neitzel *et al.*, 2021), such as progress monitoring.

These findings suggest that schools, and more particularly primary school teachers, can foster the reading achievement of at-risk pupils in their classrooms, and thus contribute to a decrease in educational inequalities. The purpose of this study is to create tools to monitor the progress in reading skills of early elementary pupils. To ensure that these tools are suitable for practice, they are created together with teachers. Furthermore, we examine how these tools meet the teachers' needs when engaging in the progress monitoring of pupils' reading skills.

1.1 Progress monitoring

Progress monitoring consists of regularly measuring and analyzing pupils' progress with the aim of adapting the instruction to their needs (Dietrichson *et al.*, 2017, 2021). This practice is at the heart of two other research strands that, despite common roots, have largely developed separately: formative assessment and data-driven decision making. Formative assessment are classroom practices where teachers and/or learners collect and interpret information about what and how pupils learn (Klute *et al.*, 2017). According to Eysink and Schildkamp (2021), formative assessment consists of five main

components: developing and sharing learning goals, collecting data about pupils' learning, identifying pupils' learning needs, acting appropriately on these in the classroom, and involving pupils in this process. Data-based Decision Making (also labeled Data-driven Decision Making) on the other hand is described as the continuous process of collecting and analyzing data about pupils' skill level in order to guide decisions on instruction (Filderman *et al.*, 2018; Schelling and Rubenstein, 2021).

Both formative assessment and Data-based Decision Making were found to be beneficial for reading skills. Indeed, Klute *et al.* (2017) report that formative assessment has a large effect on primary pupils' reading achievement (+0.41 standard deviations on achievement level compared to a control group). In addition, the meta-analysis by Filderman *et al.*, (2018) indicates that struggling readers benefit from Data-driven Decision Making (*Hedges' g* = +0.27, 95% *CI* = [0.07, 0.47]). The literature also points to the fact that students learn more when tested than when they re-study the same material. This is referred to as the "test effect" (e.g., Adesope, Trevisan and Sundararajan, 2017; Yang *et al.*, 2021). Indeed, Yang *et al.* (2021) argue that tests were not only beneficial for learning factual knowledge, but also promoted conceptual learning and facilitated problem solving. More frequent testing may therefore be another benefit of progress monitoring.

1.1.1 Collecting and analyzing data on pupils' learning

A key element of formative assessment concerns collecting data on pupils' learning (Klute *et al.*, 2017; Eysink and Schildkamp, 2021). This data can be either formal – such as exercises, homework assignments – or, informal – such as discussions with the pupil and observations of pupils while they are working on a classroom assignment (Gottheiner and Siegel, 2012; Hargreaves, 2013; Yin, Tomita and Shavelson, 2014). Data can be collected by the learner (e.g., self-assessment), or by other pupils in the class (e.g., peer-assessment, Black and Wiliam, 2009). However, Klute *et al.* (2017) have shown that for reading, it is more effective to consider formative assessment delivered by a teacher, educator, or computer program.

Data-based Decision Making is more dependent on formal data (Wayman, Jimerson and Cho, 2012; Schildkamp, 2019; Eysink and Schildkamp, 2021; Hebbeker *et al.*, 2022) and generally distinguishes between two types of measures: curriculum-based measurement (CBM) and mastery measures (Filderman and Toste, 2018; Filderman *et al.*, 2018). CBMs are short, standardized measures that indicate the overall skill level. The most frequently measure used in reading is oral reading fluency, e.g. the number of words correctly read in one minute (Van Norman, Nelson and Parker, 2018). Mastery measures are assessments of a specific skill – for example, decoding ability – and are closely linked to specific learning activities (Stecker, Lembke and Foegen, 2008). This helps to make decisions about the adjustment of the learning activity (Filderman and Toste, 2018; Van Norman, Nelson and Parker, 2018). However, as mastery measures are specific and usually not norm referenced, they do not allow for the assessment of the overall skill and comparison with average pupils are limited (Filderman and Toste, 2018). Therefore, curriculum-based measures are recommended for detecting pupils at risk, while mastery measures are recommended for more regular progress monitoring of those pupils (Van Norman, Nelson and Parker, 2018). Moreover, it is generally recommended to assess reading skills that pupils struggle with (Lemons, Kearns and Davidson, 2014).

Recommendations regarding frequency of measures of pupils' progress vary (Ardoin *et al.*, 2013). First, the baseline level of a pupil's reading skills need to be assessed, based on three measures conducted in quick succession (Lemons, Kearns and Davidson, 2014; Filderman and Toste, 2018). After this step, the goal is to gain enough data to make relevant pedagogical decisions. Depending on the stability of growth of the pupils over time (Stecker, Lembke and Foegen, 2008), some researchers

suggest a minimum of five to six weeks of data (Ardoin *et al.*, 2013) while others suggest at least 20 moments of data collection (Christ *et al.*, 2012). More precisely, Filderman and Toste (2018) recommend using more frequent data collection with struggling readers, as the deviation in their performance is larger and so the accuracy of the assessment is reduced.

After data collection, an analysis phase is needed to convert the raw data into useful information (Klute *et al.*, 2017; Eysink and Schildkamp, 2021). Data can be presented in many forms: tables, texts or graphs (Hebbecker *et al.*, 2022) When curriculum-based measurements is used, pupils' results can be compared to a pre-established cut-off score or the slope can be analyzed to establish whether progress is in line with expectations (Filderman *et al.*, 2018; Oslund, Elleman and Wallace, 2021). The aim of this analysis phase is to identify pupils' strengths and weaknesses in order to best meet their needs (Eysink and Schildkamp, 2021).

1.1.2 Differentiation to support struggling pupils

When pupils' progress is deemed insufficient, differentiation is recommended (Allal and Mottier Lopez, 2005; Klute *et al.*, 2017; Filderman *et al.*, 2018). Differentiation has been defined in many ways in the literature (Bondie, Dahnke and Zusho, 2019; van Geel *et al.*, 2019). Here, differentiation is defined as instructional adaptations in response to pupils' cognitive needs (Roy, Guay and Valois, 2013; Deunk *et al.*, 2018). It encompasses various practices such as homogenous flexible grouping or modifying learning instruction (Godor, 2021; Taylor *et al.*, 2022). Regarding the effects of differentiation on pupils' mathematics and reading performance, a recent meta-analysis concluded that there was a small, positive impact (Cohen's $d = +0.146$, 95% CI = [0.066; 0.226]) (Deunk *et al.*, 2018).

Additional evidence for the effect of progress monitoring and differentiation is presented in the studies on Response to Intervention (RTI) model (Puzio, Colby and Algeo-Nichols, 2020; Oslund, Elleman and Wallace, 2021). Various versions of the RTI model exist (Alahmari, 2019), but in its traditional form, the RTI model has three tiers, distinguished by the intensity of the support provided to the learner. Tier 1 concerns providing all pupils with the best possible evidence-based educational practices. Pupils showing difficulties are redirected to the second tier (Tier 2) where in addition to the instruction received in Tier 1, pupils receive a targeted intervention in small groups. Tier 3 is devoted to pupils with severe difficulties, which persist despite the Tier 2 intervention. These pupils receive a more intense and lengthy intervention, usually on a one-to-one basis (Greenfield *et al.*, 2010; Alahmari, 2019; Neitzel *et al.*, 2021).

Within each tier, the combination of high-quality teaching and regular assessment of pupils' skill levels ensures effectiveness (Alahmari, 2019). As such, the RTI model is based on four critical components: the presence of different tiers, screening, progress monitoring, and data-based decision making (Oslund, Elleman and Wallace, 2021). Screening, implemented in Tier 1, identifies pupils at risk (Alahmari, 2019). Monitoring the progress of these pupils makes it possible to redirect them to Tier 2 and then Tier 3 if they do not show sufficient progress (Arden *et al.*, 2017), to evaluate whether this extra support has the hoped-for results and, if not, to adjust the teaching practices of Tier 2 or Tier 3 (Alahmari, 2019).

Several studies highlight the benefits of the RTI model, while others are more skeptical. Thus, the advantages pinpointed in the literature are as follows. First, the RTI model allows for the identification of at-risk pupils as well as pupils in need of special education (Alahmari, 2019). Moreover, it enables pupils' diverse needs to be addressed and interventions to be made as soon as difficulties arise (Arden *et al.*, 2017; Alahmari, 2019). Moreover, data-based decision making brings

more benefits to pupils with difficulties, as their progress is tracked on a regular basis (using mastery measures), allowing instruction to be tailored to their needs (Oslund, Elleman and Wallace, 2021). Meta-analyses by (Slavin *et al.*, 2011) and (Neitzel *et al.*, 2021) point to the positive effects of interventions with features of the RTI model on the reading skills of struggling pupils. However, others have suggested that RTI does not work as well as expected. Balu *et al.* (2015) collected data from 20,450 first through to third graders from 146 schools in the United States. The results showed that first graders who received interventions performed statistically worse than their peers. As for those in second and third grade who received a Tier 2 or Tier 3 intervention, they did not perform better than other students. Yet, as Al Otaiba *et al.* (2019) showed the findings of Balu *et al.* (2015) should be interpreted cautiously as many schools did not consistently implement RTI using evidence-based practices (e. g. Fuchs and Fuchs, 2017; Gersten, Jayanthi and Dimino, 2017).

1.1.3 Teachers' difficulties in implementing progress monitoring and differentiation

Despite the clear benefits of progress monitoring and differentiation, these practices are underused, even in contexts that encourage their implementation (Oslund, Elleman and Wallace, 2021; Schelling and Rubenstein, 2021). Two key reasons appear to be particularly relevant. First, primary school teachers' attitudes toward the RTI model (Greenfield *et al.*, 2010; Castro-Villarreal, Rodriguez and Moore, 2014; Cowan and Maxwell, 2015) and the data-driven decision making process (Schelling and Rubenstein, 2021) are mixed. While teachers perceive the usefulness and added value of these changes (Schelling and Rubenstein, 2021), particularly in tracking progress (Greenfield *et al.*, 2010; Cowan and Maxwell, 2015), they find them a source of stress and anxiety (Schelling and Rubenstein, 2021), increasing their workload and responsibilities (Castro-Villarreal, Rodriguez and Moore, 2014; Cowan and Maxwell, 2015). Indeed, teachers indicate a lack of time and resources for implementing these practices (Castro-Villarreal, Rodriguez and Moore, 2014; Klute *et al.*, 2017). Regarding differentiation, teachers identify several factors that hamper implementation: the diversity of students in the class group, the lack of support from the school, and, the lack of rich and useful information about students' skills levels (van Geel *et al.*, 2019). Teachers in general education face these challenges in particular, as they have on average more students per class, which makes it more difficult to implement small group instruction or individual support (Alahmari, 2019).

Secondly, teachers vary in their ability to collect data, interpret it and link this information to relevant pedagogical adaptations (Greenfield *et al.*, 2010; Klute *et al.*, 2017). Primary school teachers' skills and perceived control were found to be related to their tendency to implement data-driven decision making (Prenger and Schildkamp, 2018). Nevertheless, it may be complicated for teachers to plan relevant pedagogical adaptations based on learners' needs (Colognesi and Gouin, 2022) and to measure learners' performance regularly before formal assessment (Gaitas and Alves Martins, 2017), as they are, for example, occupied with classroom management (Schelling and Rubenstein, 2021). In addition, the decision-making processes for differentiation are poorly documented in the literature (Puzio, Colby and Algeo-Nichols, 2020). For example, there are no clear recommendations on when a learner should be considered at risk and receive additional interventions (Hughes and Dexter, 2011). Some teachers report being uncertain about the boundaries between students who are expected to benefit from Tier 2 and those who are expected to benefit from Tier 3 (Greenfield *et al.*, 2010). As a result, teachers are unprepared and unwilling to implement practices such as the RTI model and its implementation fidelity is low, with high variability between schools (Arden *et al.*, 2017; Berkeley *et al.*, 2020; Oslund, Elleman and Wallace, 2021). Without this fidelity, which is also recognized as important by teachers (Greenfield *et al.*, 2010), the effects on student performance remain below expectations or are absent (Arden *et al.*, 2017; Gersten, Jayanthi and Dimino, 2017).

1.2 Changing teaching practices

According to practice-embedded research (Donovan, Snow and Daro, 2013; Snow, 2015) and improvement science (Bryk, 2015, 2017), the gap between what is recommended in research and what is done in practice (Berkeley *et al.*, 2020) is the result of research projects developing teaching programs that teachers should replicate (Bryk, 2015; Goigoux, Renaud and Roux-Baron, 2021). Thus, researchers and teachers advocate increased in-service training to improve teaching practices (Castro-Villarreal, Rodriguez and Moore, 2014; Cowan and Maxwell, 2015; Oslund, Elleman and Wallace, 2021). However, providing teachers with the latest scientific findings does not appear sufficient to bring about a change in teaching practices (Cèbe and Goigoux, 2018). Indeed, despite promising initial results, the effects of evidence-based practices may disappear when it is put into practice on a larger scale (Bianco, 2018; Bressoux, 2021).

Different hypotheses may explain this finding, such as the effects being less robust than expected or the degree to which teachers implement the program (Gersten *et al.*, 2020). For various reasons, teachers may opt not to implement (part of) the program. First, the teachers do not want to, for example, because it is too costly to implement, it goes against their own experience or their own beliefs and conceptions regarding teaching and learning (Caena, 2011; Bressoux, 2021; Hanin *et al.*, 2022), or they feel that their usual practices are no worse than what is being proposed and thus have little motivation to change (Quinn and Kim, 2017). Second, sometimes teachers cannot implement the program, because of external constraints, such as their school principal's visions or the attitude of parents, who might, for example, see differentiated instruction as unequal treatment of pupils (Coppe *et al.*, 2018). They may also believe that these new practices are not easily applicable to all classrooms, in all contexts (Quinn and Kim, 2017). Third, teachers may lack the necessary didactic skills and experience or else adequate implementation is difficult due to the absence of clear instructions (Bressoux, 2021). In other words, the implementation is hampered by the teaching program being poorly adapted to future users and the complex environments in which they work (Bryk, 2015).

Based on these findings, practice-embedded research aims to minimize the distance between researchers and teachers from the very beginning of the research process (Snow, 2015). There are two prerequisites: the need to consider the complex environment in which teachers practice and sustainable partnerships between researchers and practitioners (Snow, 2015). Indeed, the classroom is an environment that has become increasingly complex over time: teachers face more heterogeneous groups and interact with a variety of professionals, such as speech therapists or school psychologists (Bryk, 2015). To ensure that a program or tool is useful for practice, this complexity of the school environment needs to be integrated into the design process from the beginning (Class and Schneider, 2013). According to Snow (2015), to improve teaching practice, practitioners' experience is just as valid a source of knowledge as scientific theories. The diversity of settings in which a teaching practice or tool is more or less effective allows the identification of the necessary conditions for the implementation (Class and Schneider, 2013) and may even provide an opportunity to explore factors for improvement (Bryk, 2015). So, rather than being constraints or variables that need to be controlled, the complexity of practice settings provides essential information (Snow, 2015). Thus, the use of the teaching practice in increasingly diverse settings allows for more insight into its effect and conditions (Bianco, 2018).

The second prerequisite concerns sustainable partnerships between teachers and researchers, which is indispensable for the latter to have access to the complex reality of teaching practice (Donovan, Snow and Daro, 2013). Thus, practice-embedded research promotes collaborative research in which

researchers and practitioners are on an equal footing (Snow, 2015; Goigoux, 2017). Partnering with teachers from the start also allows for the development of tools that consider future users' needs (Cèbe and Goigoux, 2018). Furthermore, it is essential that the tools or programs fit into the existing habits of the teachers (Goigoux, Renaud and Roux-Baron, 2021).

Thus, rather than waiting for teachers to adapt their practice to researchers' recommendations, the objective is to construct a program or a tool together, adapted to practitioners' needs (Class and Schneider, 2013; Goigoux, Renaud and Roux-Baron, 2021). For Goigoux, Renaud and Roux-Baron (2021), the quality of design predicts its acceptability among teachers and thus its implementation fidelity. Therefore, a properly designed tool should not require additional support from researchers when it is implemented.

1.3 The present study

To reduce the reading achievement gap, progress monitoring appears promising (Dietrichson *et al.*, 2017). Yet, teachers consider this pedagogical practice difficult to implement (Castro-Villarreal, Rodriguez and Moore, 2014) and recommendations from research only slowly make their way to practice (Berkeley *et al.*, 2020). To create tools suitable for practice, practice-embedded research suggests considering the complexity of the teaching practice from the onset and building sustainable partnerships between teachers and researchers (Bryk, 2015; Snow, 2015; Goigoux, 2017).

Thus, we followed a group of elementary school teachers for 4 months. Using an iterative and participatory method, they co-created, with a reference researcher, tools to monitor the progress of early elementary school pupils in reading. Qualitative analysis of all discussions on this co-creation allows us to answer the following research question: how do the tools meet the needs of primary school teachers to monitor the progress of their pupils' reading skills?

2 Method

The method follows the recommendation of practice-embedded research (Snow, 2015) and educational design research (The Design-Based Research Collective, 2003). As advocated by (Goigoux, 2017; Cèbe and Goigoux, 2018) and used in similar research (e. g. Bogaerds-Hazenberg, Evers-Vermeul and Bergh, 2019), we chose an iterative participatory process where a group of volunteer teachers work together to improve a prototype tool during focus groups. The prototype is then tested by the teachers in their classrooms and they bring their suggestions for improvement to the next meeting. These suggestions foster the development and improvement of the tool. Disagreements, either between teachers or between teachers and researchers, are especially discussed in order to work towards a common creation, both acceptable and feasible for all teachers and respecting the initial objectives of researchers.

2.1 Context of the study

As respective policies encouraged these practices, the Response to Intervention model is mainly present in the United States (Neitzel *et al.*, 2021) and Data-based decision making is more studied in the Netherlands (Visscher, 2021). Our study takes place in the context of the French-speaking part of Belgium where such government recommendations do not exist. Furthermore, despite external tests and the dissemination of teaching guidelines, each teacher has a great degree of freedom to achieve the expectations set by the school curriculum (Dupriez, 1999; Renard *et al.*, 2022). In this way, each teacher can select the textbooks, tools, materials, etc. that they use in their classroom.

2.2 Participants: the group of co-developers

The group of co-developers is composed of a reference researcher and five first and/or second-grade teachers from different schools. Table 1 provides an overview of their main characteristics. In addition to their teaching degree, all of them pursued or are pursuing other qualifications such as a Master's in educational sciences. Furthermore, three teachers out of five have additional experience in teaching reading (e.g., co-authors of a teacher's manual for reading and writing, and participation in a field experiment on the effect of co-teaching on pupils' reading performance). The socio-economical background of their pupils ranged from highly disadvantaged to strongly advantaged.

Table 1.

2.3 Procedure: focus group meetings between co-developers

The group of co-developers met four times between September 2021 and December 2021. The reference researcher¹ organized the meetings and moderated the discussions. The first two meetings took place in Carol's and Sophia's classrooms respectively. Due to the evolution of the COVID-19 pandemic, the last two focus groups were organized online. Table 2 provides an overview of the four focus groups and their main objectives.

The objectives of the first focus group were threefold. In the first place, the participants got to know each other and the guidelines for the collaborative work were agreed upon. Three aspects were discussed, as recommended by Van Nieuwenhoven and Colognesi (2015). First, the group dynamic: This involves aiming for a symmetrical relationship between the individuals, so that everyone would dare to share and could contribute according to their expertise. Second, the usefulness of collaborative work: what the group can bring to its members and to the teaching community as a whole. Third, the organizational aspects, involving discussing spatial (where the meetings take place) and temporal constraints (what are the most appropriate times for members, how to ensure that everyone is available). Then, a brief theoretical explanation of progress monitoring and the expected positive effects on pupils' learning was provided. Mary, a doctoral student specializing in prerequisites for learning to read was present, as an additional resource to introduce the theoretical background about learning to read. Finally, the reference researchers presented three tools available in the literature to measure pupils' reading levels at the beginning of primary school: the assessments designed within the "PARLER" program (Zorman *et al.*, 2015), the tool for identifying learning outcomes in reading in first grade ("OURA LEC/CP", Billard *et al.*, 2013), as well as the assessment sheets proposed in the "Reading Workshops" (Calkins, 2017). These tools include assessments of prerequisites and components of reading, such as phonological awareness, decoding, comprehension, fluency, and the concept of print. Teachers were also invited to bring in any useful resources. During the first focus group, participants provided consent to record future focus groups.

Between the first and second meetings, participants were invited to examine in more detail the tools brought by the reference researcher. Victoria also shared an extra resource from the government (Deum *et al.*, 2007) and individual assessment sheets that she uses with her pupils.

During the second focus group, the group of co-developers work together to create the first prototype of the criterion-based rubric which aims to assess the skill level of a single pupil. They developed

¹ The reference researcher is the first author of the article.

two versions: a landscape one, which allows multiple evaluations – and thus assesses the progress over time – and a portrait one, with a comment section.

Between the second and third focus groups, teachers were invited to use the tools in their classrooms. In addition, Sophia gave the tool to her colleague to get an external point of view. Based on these experiences, during the third focus group, the landscape version of the tool was seen as more useful. Hence, the portrait version was dropped. For the landscape version, some criteria were also adjusted. Furthermore, they developed a whole-class tool with the same criteria as the individual criterion-based rubric and a whole-class tool for letter-sound correspondence.

Again, between the third and the fourth meeting, teachers were invited to test the tools in their classrooms and Carol gave them to a colleague, in order to obtain additional suggestions. Based on their feedback, the group of developers discussed adaptations of the built tools. They also decided to provide a blank version of the whole class tool for letter-sound correspondence. Furthermore, they wrote the appendix for future users, mainly based on the suggestions of the two users who did not participate in the collaboration process. Finally, the group of co-developers shared their perceptions of the processes of collaboration.

Table 2.

2.4 Data analysis

To analyze how the developed tools met teachers' needs when monitoring their pupils' progress, we first transcribed more than ten hours of audio-recorded discussions between the co-developers (for a total of 257 pages). Then, the focus group transcriptions were analyzed following the procedure described by (Baribeau, 2009). In the initial phase, the first author performed a first, inductive coding using the software Taguette. Next, the three authors discussed these codes. Based on the results of this discussion, the first author coded the transcripts anew to refine the coding and to categorize the codes into the main needs, which were discussed anew by the three authors.

Given that decisions made on the tools are the results of discussions between the developers, an interactionist framework was also chosen. The selected framework separates three areas of knowledge: shared, accepted, and disputed (Morrisette, 2011a; Morrisette and Guignon, 2016). Shared knowledge characterizes points of discussion on which participants agree. Accepted knowledge represents whatever received neither the absolute approval nor disapproval of participants. Disputed knowledge is the result of strong disagreements among the developers. Although this analytical framework was initially created to classify discussions between teachers about their teaching practices (Morrisette, 2011b), it has also been used in various contexts, generally in group interviews (e.g., Nadeau, 2021). With regard to the progress monitoring in reading at the beginning of elementary school, this framework highlights both what is shared, and therefore the professional routines into which the tools must be integrated, what is accepted, which may constitute avenues for improving teaching practices, and what is disputed, representing a probable obstacle for dissemination.

3 Results

3.1 Tools for progress monitoring

In line with the aim of developing tools to monitor first-grade pupils' progress in reading, participating teachers and a researcher co-created tools that were improved as the focus groups

progressed, based on teacher feedback. Thus, the co-creation process resulted in four tools (see Supplementary material). The first tool is a criterion-based rubric for reading components, targeting the progress of one pupil over time. It contains 28 criteria grouped into 5 categories: phonological awareness, rapid naming, phonics, global reading of function words, and understanding. The second tool is the whole-class version, allowing a teacher to assess all pupils in the classroom using one document. The third tool is a whole-class tool for letter/sound correspondence. Finally, a blank version is also suggested. In addition, the developers have written an appendix containing the instructions for use and some theoretical details.

3.2 Teachers' needs to monitor progress in reading

To answer the research question on teachers' needs when monitoring pupils' progress, the analysis of discussions between the participating teachers revealed four important needs to be met by the constructed tools: perceived usefulness, limiting the workload, balancing workload and perceived usefulness, and flexibility. In line with the interactionist framework, these can be grouped into three areas of knowledge: shared, accepted, and disputed. Table 3 provides an overview of these results, which are further detailed below.

Table 3.

3.2.1 Perceived usefulness

3.2.1.1 Shared area

The co-developers attempt to optimize the perceived usefulness of the tools by multiplying the objectives facilitated by them. Thus, participating teachers share three common goals: identify the level differences between pupils, differentiate and log information.

The tools allow for the observations of pupils' progress over time, as they evidence the evolution of the number of criteria met. *"And so, if we want to monitor progress, [...] we have to be able to situate the pupil"* (Sophia, Focus group 1, further abbreviated as FG1). Teachers who use the tools feel that the differences between pupils were identified and biases were avoided. As Mary points out (FG 1), *"The risk in [...] not identifying a pupils' level is that sometimes [...] we put him in a category by saying to oneself [...] it's not going okay. Whereas, if we had, if we had assessed the pupil's level individually [...] he might have been quite successful."*² George notes the opposite risk: *"And more seriously, the opposite is also true. Because ... for example, a super shy pupil, I realized in December that he was struggling a lot while ... [for] me, it was going okay"* (FG 1).

Determining whether pupils master certain skills allows those at risk to be identified. Furthermore, it helps to identify which specific skills to target during differentiated instruction, which can take the form of for example additional exercises or additional time with the teacher alone or in small groups. Mary explains: *"Here we can tick off, if we didn't check off [the criteria], that means that these pupils must be worked with separately. But the others keep going and the pupils who have not yet acquired the skills keep going, but we thus saw it in time and we remedy in time. Lucy: That's it".* The goal is to identify struggling pupils in time to offer them differentiated instruction, and so, prevent a widening of the gap between high and low performers.

² The verbatims were translated from French.

Finally, the tool makes it possible to log all of the teachers' reflections when they observe their pupils or analyze the results of a test. These notes indicate whether an error persists and can be used as a support for communication with parents, speech therapists, or other teachers. In addition, teachers can communicate this information to the pupils, using verbal or written feedback. The tools also make it possible to assess the effects of differentiated instruction.

“Researcher: And like this, 3 weeks later, I return to this tool for this pupil and say to myself: oh right, 3 weeks ago, I had set up this, and well, now the pupil masters it.

Carol: And that makes the teacher to go that far in his reflection and to say to himself, ah yes, I've observed that, because sometimes we observe things.

Sophia: And we stop there [...]

Lucy: Yes, absolutely. Often due to a lack of time, so we overlook

Carol: While it's essential to get to that point” (FG 2)

3.2.1.2 Accepted area

Having a list of the essential steps a pupil has to go through to learn to read is considered useful, primarily for novice teachers. Indeed, the tools summarize the essential steps in learning to read. From prerequisites such as phonological awareness to decoding and comprehension of sentences and texts, these tools allow participating teachers to have clear criteria to assess their pupils' level of proficiency. As Carol explains, *“Here, I have the feeling that it is written in a very concrete way, as the teacher would put it into practice in their class” (FG 4)*. This aspect is mainly considered useful for novice teachers, as it shows them a way to structure their observations and, like this, evaluate all the relevant criteria. As George states (FG 3): *“Well, it allowed us to really have an entire summary and to think about all the aspects.”*. However, in order to keep the use of the tools for reading components manageable, they do not aim to be comprehensive. Indeed, George explains (FG 4): *“There are limits to these tools, it should not be taken as a bible but rather as tools allowing us to test certain aspects that we felt were important in learning to read.”*

In addition, the different tools allow participating teachers to accomplish different, varying goals, but whose complementary nature is recognized. On the one hand, the whole-class tool for reading components facilitates creating homogeneous ability groups: pupils with the same difficulties can be easily identified and grouped for differentiated instruction. On the other hand, the rubric for reading components is more precise and can be integrated into a tailored learning path. The variety of tools allows future users to select the tool according to their needs. Lucy explains (FG 3): *“I think that [...] the two versions can complement each other as you said... Well, yes, why not for struggling pupils keep the individual one [...] And as it was said to create ability groups, having the collective one can be interesting too.”*.

3.2.2 Limiting the workload

3.2.2.1 Shared area

Participating teachers perceive progress monitoring as a time-consuming practice. Indeed, collecting data and analyzing it takes a lot of time, especially if teachers have to assess each pupil individually.

“Lucy: Often, we don't take the time to analyze them [the errors], that's the problem.

Carol: Because it would take too much time.

461 *Lucy: Because we don't have the time" (FG 2)*

462 Therefore, the co-developers set out to build tools that can be used with minimal time and effort. To
463 do so, they opted a format that they considered easy to use (i.e., few columns and in landscape
464 format). The developers also opted to write an appendix (see Supplementary material), which allows
465 some theoretical points to be clarified. However, participating teachers insisted on a concise
466 appendix: it should not exceed a few pages. In addition, by relying on the vocabulary in use, which
467 allows an intuitive understanding of the criteria, their goal was to match teachers' routines. The
468 criteria were arranged chronologically, according to the usual sequence in which a pupil learns to
469 read, and classified under explicit titles. Furthermore, the criteria were intended to be easily
470 observable. The goal of all these measures was to reduce the time and effort required for adequate
471 use of the tools.

472 *"Lucy: It should not be ...*

473 *Carol: Time-consuming*

474 *Lucy: Yes, that's it, time-consuming and that we have to, I think about my multi-age*
475 *classroom here, my 21 pupils, I'm thinking, if I have to call them one by one while ensuring*
476 *that the others are not lost.*

477 *Victoria: That's it." (FG 1)*

478

479 **3.2.3 Disputed area**

480 The participating teachers discussed at length balancing workload and perceived usefulness. During
481 this discussion, two groups of teachers emerged. The first group considered using the tools too time-
482 consuming compared to the benefits. They considered the rubrics for reading components to be too
483 precise, as they contain various criteria, thus increasing the time investment. In addition, they said
484 that the tools do not provide enough additional elements compared to their usual practices. As Carol
485 explained: *"I have just finished some evaluations and in fact, when I was done, I told myself: I did*
486 *not use the tool. And I read the tool, I thought, but actually, I just did all the work for the school*
487 *reports [...]. This tool isn't actually going to help me."* (FG 3). For these teachers, the progress
488 monitoring tools are similar to the school report. Victoria said (FG 3): *"But it's because for me,*
489 *presented in an individual way, it's similar, it's similar to my skills report, really"*. However, if the
490 rubric for reading components is used as a school report, the effort of filling out this rubric must be
491 made for each pupil, which considerably increases the workload. Consequently, they intend to use
492 the tools less frequently, on average four times a year.

493 The second group disagrees with the goals of the rubrics. First, these teachers perceive the tool to be
494 too precise for a school report, and it contains too many technical terms, which could hamper
495 communication with parents. Secondly, they mainly use the rubrics for reading components for
496 struggling pupils. As Georges says, *"I really agree with you [about the workload required] but I*
497 *didn't use it for the children who read easily but [...] for the children with difficulties, [...] it allowed*
498 *me to find all the aspects that were still complicated for them."* (FG 3). Thirdly, they plan to assess
499 the progress of their struggling pupils on a regular basis, and more frequently than their peers.

500 These two profiles differ in the number of years of experience in teaching and their class
501 organization. Indeed, the participating teachers belonging to the first group have more years of
502 experience (18 and 24 years) compared to the second group (7 to 9 years). As Sophia explains (FG
503 3): *"And so, when we saw this tool, we said wow, so good, finally something that will structure our*

504 *thoughts, our work and everything. And when I hear Carol who has already explained to us a little*
505 *bit about her way of working and everything, Carol you are someone, it seems to me, [...] much more*
506 *organized. [...] Your way of evaluating is well planned, step by step, et cetera. And so, I guess, I can*
507 *understand that in fact, there are like two types of people.”*

508 **3.2.4 Flexibility**

509 **3.2.4.1 Shared area**

510 To optimize the acceptability of the tools among the teaching routines, the co-developers decided to
511 provide future users with a large number of options. As Carol summarizes:

512 *“Anyway, if we try to impose something on teachers, they will do as they please. Let’s be*
513 *honest!*

514 *George: That’s true” (FG2).*

515 Consequently, the developers aimed to create flexible tools: the co-developers wanted that the tools
516 could be easily adjusted by the teacher, in line with their pedagogical practices, program, or method
517 for teaching reading. In addition, they did not impose a frequency for using the tools, so it could vary
518 according to the teacher’s preference.

519 *“Sophia: In any case, there are simple and complex [sounds], as a comment the teacher can*
520 *put, “oi, aw, oo” and “dle, gle”...*

521 *Carol: He does as he sees fit!*

522 *Sophia: That’s it*

523 *Carol: Anyway, you’ll do as you please, no?*

524 *Sophia and Lucy: Well, yes.*

525 *Sophia: And if the teacher can’t make it his own, he won’t do it, he won’t use it anyways*

526 *Lucy: That’s true” (FG 2)*

527 The “blank” version of the tool is further proof of the pursuit of adjustable tools. This version is very
528 simple and allows teachers to choose the order in which they want to evaluate letter-sound
529 correspondences, in line with their program and teaching routines.

530 The expertise of users is valued as well. The co-developers consider teachers to be competent to
531 provide relevant interventions to remedy pupils’ difficulties and to hypothesize explanations for their
532 origins. The large boxes in the rubrics for reading components allow teachers to provide comments
533 on their observations. In addition, the developers perceive teachers as sufficiently qualified to decide
534 how to assess pupils’ progress. Thus, the tools can be completed on the basis of formal, informal,
535 individual, and/or group assessments. Indeed, the participating teachers consider that, depending on
536 the criteria, the teachers’ routines, and experience, it was impossible to provide the same formal
537 evaluation sheet for all users.

538 *“Researcher: But how am I going to evaluate this? Do I approach the child and have him*
539 *read 10 words and do it like that or actually, I see the child all day ...*

540 *Lucy: I observed [...]. I would like to say that it depends on [...] [the] skills and there are*
541 *some that can easily be observed and others. So, if we want to evaluate oral reading fluency*
542 *at some point, it should be good to assess them individually.*

543 *Sophia: To let them come see me, yes that's it. [...]*

544 *Carol: Yes, but [...] in first grade, you make them read every day or you try. You can quickly,*

545 *Lucy: Yes, you can quickly complete it, that's true. (FG 2)*

546 Therefore, filled-out tools cannot be compared between pupils, especially if they are in different

547 classes. For the developers, the main goal is to help teachers monitor their pupils' progress in reading

548 in first grade. Hence, inter-individual or inter-class comparisons are considered less relevant.

549 In addition, the tools can be adjusted to the pupils, their level, and their needs. The teacher is not

550 required to complete all of the criteria to make decisions on adapting their teaching practices.

551 Depending on the time of year or the pupils' abilities, some items may be unnecessary and redundant.

552 *"Carol: There are pupils who need to*

553 *Victoria: to go through the intermediate phase.*

554 *Carol: That's right, but it doesn't concern all of them. Well, me ...*

555 *Victoria: Let's say that it can be an extra criterion.*

556 *Carol: But again, you can put it as a special criterion. Afterward, it's up to the teachers to*

557 *see if they complete this criterion or not." (FG 3)*

558 Therefore, the length of the tools and the resulting workload vary depending on the pupils and the

559 contents already taught.

560 **3.2.4.2 Accepted area**

561 The participants discussed the format of the evaluation as well. Some of the developers argue for a

562 grade, which they believe to be more accurate and objective. Others feel that quantitative assessments

563 are just as subjective as qualitative ratings, but if marks were communicated to pupils, the class

564 atmosphere could be hampered. All criticize a fixed threshold of "50%" for success. On the one hand,

565 as Carol explains: *"You can't read when you can read one word out of two" (FG 2)*. According to

566 the participating teachers, it is necessary to keep helping the pupils until they master the targeted

567 skill, which corresponds to a grade well above 50%. On the other hand, they note that a fixed

568 threshold of 50%, based on a single assessment, entails serious risks: a pupil with 51% would not

569 receive the same support as a pupil with 49%, even though they both need it.

570 For these reasons, the co-developers agreed on leaving the format of the evaluation to the teachers,

571 but recommending the categories "Acquired – Not Acquired". This ensures a usable format for

572 everyone, regardless of whether the completion of the tool was the result of a classroom observation

573 or a formal assessment. Some teachers, like Sophia, argue for the addition of an "In the process of

574 acquiring" category. She was also more comfortable with the use of a more precise percentage, to

575 help her quantify how little or how much a skill was acquired.

576 *"For me, a pupil where it's 'not acquired' but it's not acquired at 45%, there's only a small*

577 *step, but not acquired where we see that we're in the 10-15%, the step is going to be huge,*

578 *there will be more work to do." (FG 2).*

579 However, others argued that the addition of another category makes the boundaries between each

580 category more subjective. In addition, since reading is intensively trained in first grade, a pupil's

581 progress is seen in the number of criteria or letter-sound correspondence acquired over time rather

582 than within each criterion. Thus, they feel that the use of a third category or a percentage were more

confusing than helping and that the precision was unnecessary. Despite different practices, the participating teachers agreed on a flexible categorization with the possibility of adding comments. In addition, it allows teachers to distinguish two groups easily: pupils who have mastered the given skill and those who need additional support.

4 Discussion

Progress monitoring has been highlighted as a fruitful teaching practice to reduce the reading achievement gap (Dietrichson *et al.*, 2017; Klute *et al.*, 2017). Yet, progress monitoring is only rarely used in practice, as teachers find it cumbersome to implement (Castro-Villarreal, Rodriguez and Moore, 2014; Cowan and Maxwell, 2015). To create tools suitable for practice, the present study relied on practice-embedded research, based on an iterative and participatory process involving five teachers. This resulted in four tools to monitor pupils' progress in learning to read at the start of primary education.

Content analyses of the discussions between the developers using an interactionist framework (Morrisette, 2011b; Morrisette and Guignon, 2016) revealed three shared areas of knowledge: perceived usefulness, flexibility, and limiting the workload. At first sight, these needs closely resemble the dimensions as put forward in the Continuous Use Design (Renaud, 2020): usefulness, usability, and acceptability. Indeed, the first dimension includes the relevance of the objectives of the devices, which is similar to the perceived usefulness of our results. The second dimension, usability, can be linked to developers' desire to limit the workload and optimize the flexibility of the tools, particularly in relation to the target group of pupils. Finally, acceptability in the continuous use design focuses on the compatibility between the tool and the characteristics of the teacher, such as their values and pedagogical style. However, according to the developers in the present study, this acceptability depends on the tools' flexibility: to guarantee the integration of a tool into the teaching habits, it needs to be easily adjustable to teachers' practice and to pupils' level and needs. The results of the present study also further refine the acceptability dimension as put forward by Renaud (2020): rather than considering the three dimensions separately, the tendency to use the tools was found to depend on the balance between the perceived usefulness on the one hand and the workload that a tool requires on the other.

In line with the need for perceived usefulness, the developers agreed that the tools allowed them to identify pupils' level differences, log information, and differentiate according to pupils' needs. These findings resemble the key ideas of the Response to Intervention model: the tools allow them to identify those who are floundering and offer them additional support, which corresponds to Tier 2 of the model (Alahmari, 2019). In addition, in line with recommendations (Arden *et al.*, 2017; Filderman and Toste, 2018), some of the teachers used the tool as a more detailed and regular follow-up for struggling pupils. Furthermore, it allows teachers to put pupils with the same difficulty together in homogeneous ability groups (Puzio, Colby and Algeo-Nichols, 2020). Yet, if these groups persist over time, the effect of this form of differentiation can be disadvantageous for struggling pupils (Deunk *et al.*, 2018), possibly even increasing the achievement gap. This is in line with (Denessen, 2017) on the risk of a possible divergent effect of differentiation as teachers may offer fewer learning opportunities to struggling pupils.

Yet, still in light of the perceived usefulness, developers accepted that the tools summarize essential steps for learning to read and that different tools may serve different goals. These goals are similar to those of the progress monitoring literature: the tools constructed allow teachers to facilitate data collection on pupils' mastery levels, to provide feedback to pupils, and to translate this information

into actions targeting struggling pupils, in the form of differentiation. These steps are also (in part) identified in the literature on progress monitoring (Dietrichson *et al.*, 2017), formative assessment (Klute *et al.*, 2017), and data-driven decision making (Filderman *et al.*, 2018).

The developers also shared a clear desire to limit the workload, as they perceived progress monitoring as cumbersome and time-consuming. These perceptions are consistent with the literature on teachers' attitudes to data-driven decision making (Schelling and Rubenstein, 2021) and the Response to Intervention model (Greenfield *et al.*, 2010; Cowan and Maxwell, 2015). To this end, the developers ensured that the format of the tools facilitated easy use. However, contrary to what teachers in the context of the USA have suggested (Castro-Villarreal, Rodriguez and Moore, 2014; Schelling and Rubenstein, 2021), the help of colleagues, as well as the prospect of additional resources provided by the school, were not mentioned as strategies to decrease teacher workload. Possibly, this is linked to the lower general level of teacher collaboration: in the French-speaking part of Belgium, teacher collaboration was found to be below the OECD mean (Quittre, Dupont and Lafontaine, 2021).

Despite a common desire to reduce the workload and optimize perceived usefulness, two groups emerged when developers balanced both needs. The first group of teachers saw the rubrics for reading components as a report card, necessary for all pupils. The second group used this tool primarily for struggling pupils. The developers identified two characteristics in teachers that set both groups apart: the degree to which a teacher is well-organized and years of seniority. Well-organized and more experienced teachers tended to belong to the first profile. It is possible that their positions are also influenced by their conceptions of justice. Indeed, Vijfeijken *et al.* (2021) examined teachers' arguments to justify their differentiation practices and classified them using the principles of distributive justice: equality (i.e., an equal distribution of resources and/or the same expectations for all learners), equity (i.e., a distribution of resources proportional to their merit such as effort by the learner) and needs (i.e., an unequal distribution of resources based on learners' needs). The research by Vijfeijken and colleagues found that these principles of justice were linked to teachers' practices of differentiation. Thus, it is possible that, in the present study, the developers who wish to devote more time and effort to struggling pupils (group 2) justify – unconsciously or not – their practices with principles based on learners' needs, and that the teachers in the first profile place more emphasis on equality. Hence, future research on teachers' use of progress monitoring tools could specifically examine these principles of distributive justice.

The developers also agreed on the need for flexibility. Consequently, the tools are easily adjustable to pupils' levels and needs and to teachers' preferences for monitoring progress. This corroborates the finding by Schelling and Rubenstein (2021) that teachers generally prefer to use their own assessments over standardized tests. Moreover, Van der Kleij *et al.* (2015) found that foster teachers' sense of autonomy is linked to a successful implementation of Data-based decision making. However, as pointed out by the co-developers, this has the consequence of limiting comparisons between teachers.

Furthermore, although the developers' initial goal was to reveal pupils' skill level differences, one may wonder whether this flexibility may – unintentionally – provide more room for these biases to impact teachers' judgement. Indeed, teachers tend to have lower expectations for students from disadvantaged backgrounds (for a review, see Wang, Rubie-Davies and Meissel, 2018) and multiple studies have found teacher bias in assessments regarding pupils' background (Hanna and Linden, 2009; Sprietsma, 2013; von Hippel and Cañedo, 2022). A recent literature review has shown that teachers' implicit biases sometimes predict their behavior better than their explicit attitudes

(Denessen *et al.*, 2022). For example, Gortazar, Martinez de Lafuente and Vega-Bayo (2022) conducted a large study comparing the grades awarded to an assessment by two raters: an external assessor and pupils' primary school teachers. For languages (Basque and Spanish), results indicate that boys, first and second-generation immigrants, and pupils from disadvantaged backgrounds are judged more negatively by their teacher than by the external assessor. These biases may play a stronger role in teachers' judgement when tools are flexible. Indeed, the study by Quinn (2020) found that, when using a detailed rubric (implying a low level of flexibility), this led to a fairer judgement of the skill level of ethnic minority pupils. In other words, the flexibility could lead to a disadvantageous assessment of pupils from disadvantaged backgrounds and, when not combined with differentiated support, increase educational inequalities.

Developers agreed that the practice field is complex and diverse, as previously highlighted in the literature on practice-embedded research (Snow, 2015; Goigoux, Renaud and Roux-Baron, 2021). Consequently, diverging views were considered inevitable and the need to value teacher expertise was underlined. In this way, the complexity of the field was handled through the flexibility of the tools. This flexibility extends the conditions under which the tools can be implemented, which is advocated in practice-embedded research (Class and Schneider, 2013).

In light of the need for flexibility, developers have deliberately allowed teachers liberty in the format of the evaluation (accepted area of knowledge). This implies that a very wide range of information sources can be considered, such as formal assessments and classroom observations, both for one specific, struggling pupil (tool 1) or for the entire classroom (tools 2 & 3). Within these tools, qualitative and quantitative data are considered equal sources of information. This position is contrary to the literature on data-based decision making, which advocates quantitative, even standardized data (Filderman *et al.*, 2018), but is closer to formative evaluation, which defines the term 'data' more broadly (Allal and Mottier Lopez, 2005; Eysink and Schildkamp, 2021).

4.1 Limitations and implications for future research

Some limitations of the present study need to be acknowledged. First, the tools were constructed by a small group of volunteer teachers, who all pursued or are currently pursuing additional qualifications such as a Master's in educational sciences. In addition, the views of the teachers evolved over the course of the different focus groups. Although the developers attempted to create the most flexible tools for any type of primary education context, due to the limited sample and the impact of the joint creation process, future research should examine whether teachers without master level training and who did not participate in the focus group discussions can easily use them. There is early, anecdotal evidence that this is possible: two teachers gave the tools to a colleague, who found them useful. It is clear that more experiences from teachers who did not participate in the development are welcome, as these may further refine the tools (Cèbe and Goigoux, 2018). This may also point to other key teacher characteristics, besides seniority and the degree to which one is organized as detected in the present study.

Second, it needs to be emphasized that the tools were developed in the context of learning to read French at the start of primary education. Reading is a complex skill and multiple components interact when learning to read (Scarborough, 2005; Peters, Hebbeker and Souvignier, 2022). This complexity prompted the developers to create flexible progress monitoring tools. It remains to be investigated whether progress monitoring tools for reading in the later years of primary education or for other key content domains (e.g., mathematics) require the same level of flexibility.

Third, although progress monitoring and, more broadly, formative assessment are believed to foster pupils' achievement (Dietrichson *et al.*, 2017; Klute *et al.*, 2017), the present study did not set out to examine whether the co-created tools live up to this claim. Further research is needed on whether progress monitoring using these tools positively impacts pupils' reading achievement and if the tools differ in this respect. For researchers examining educational inequalities, this is also timely as all previous studies on reading combined progress monitoring with other teaching practices aimed at reducing the achievement gap (Dietrichson *et al.*, 2017, 2021). Hence, the precise effect of progress monitoring in itself remains unclear. To design adequate interventions combining multiple teacher practices to reduce the achievement gap, it is first important to gain an insight into the effectiveness of each teacher practice separately.

Finally, it is worth emphasizing that while progress monitoring may have a positive effect, it is unlikely that this practice reduces the achievement gap to an acceptably low level. Rather, it is likely to be a necessary first step in identifying struggling pupils and providing adequate interventions for these pupils. While the co-developers in the present study were confident of their own ability as teachers and that of their colleagues to provide relevant interventions, this merits further research as well.

4.2 Implications for practice

The present research expands on the previous literature with regard to teaching practices targeting a decrease in educational inequalities and the literature on progress monitoring more specifically. Rather than a researcher-led development of progress monitoring tools, the present study relied on a practice-embedded research: teachers and researchers co-created tools to help monitor pupils' progress in reading. This resulted in four tools (see Supplementary material) that practitioners can use. In addition, the tools may become part of the resources provided during teacher training.

Moreover, the content analysis of the focus group discussions revealed an important topic for future professional development. The developers discussed at length balancing workload and perceived usefulness. If schools want to put progress monitoring in place, it is likely that a disputed area of knowledge would cause disagreement among teachers. Hence, professional development in school teams could anticipate and ensure that teachers can express their views and that a consensus can be reached on this topic.

5 References

Adesope, O.O., Trevisan, D.A. and Sundararajan, N. (2017) 'Rethinking the Use of Tests: A Meta-Analysis of Practice Testing', *Review of Educational Research*, 87(3), pp. 659–701. Available at: <https://doi.org/10.3102/0034654316689306>.

Al Otaiba, S. *et al.* (2019) 'Elementary teacher's knowledge of response to intervention implementation: a preliminary factor analysis', *Annals of Dyslexia*, 69(1), pp. 34–53. Available at: <https://doi.org/10.1007/s11881-018-00171-5>.

Alahmari, A. (2019) 'A Review and Synthesis of the Response to Intervention (RtI) Literature: Teachers Implementations and Perceptions', *Journal of Education and Practice*, 10(15), p. 8. Available at: <https://doi.org/10.7176/JEP/10-15-02>.

Allal, L. and Mottier Lopez, L. (2005) *Formative assessment: improving learning in secondary classrooms*. Edited by Centre for Educational Research and Innovation. Paris: OECD.

- 756 Arden, S.V. *et al.* (2017) ‘Toward More Effective Tiered Systems: Lessons From National
757 Implementation Efforts’, *Exceptional Children*, 83(3), pp. 269–280. Available at:
758 <https://doi.org/10.1177/0014402917693565>.
- 759 Ardoin, S.P. *et al.* (2013) ‘A systematic review and summarization of the recommendations and
760 research surrounding Curriculum-Based Measurement of oral reading fluency (CBM-R) decision
761 rules’, *Journal of School Psychology*, 51(1), pp. 1–18. Available at:
762 <https://doi.org/10.1016/j.jsp.2012.09.004>.
- 763 Balu, R. *et al.* (2015) *Evaluation of Response to Intervention Practices for Elementary School*
764 *Reading*. NCEE-2016-4000. National Center For Education Evaluation and Regional Assistance, p.
765 308.
- 766 Baribeau, C. (2009) ‘Analyse des données des entretiens de groupe’, *Recherches qualitatives*, 28(1),
767 p. 133. Available at: <https://doi.org/10.7202/1085324ar>.
- 768 Berkeley, S. *et al.* (2020) ‘A Snapshot of RTI Implementation a Decade Later: New Picture, Same
769 Story’, *Journal of Learning Disabilities*, 53(5), pp. 332–342. Available at:
770 <https://doi.org/10.1177/0022219420915867>.
- 771 Bianco, M. *et al.* (2012) ‘Impact of early code-skill and oral-comprehension training on reading
772 achievement in first grade’, *Journal of Research in Reading*, 35(4), pp. 427–455. Available at:
773 <https://doi.org/10.1111/j.1467-9817.2010.01479.x>.
- 774 Bianco, M. (2018) ‘La réponse à des questions cruciales en éducation réside-t-elle dans un
775 changement de paradigme ?’, *Éducation et didactique*, (12–1), pp. 121–128. Available at:
776 <https://doi.org/10.4000/educationdidactique.3111>.
- 777 Billard, C. *et al.* (2013) ‘OURA LEC/CP Outil enseignant’. Available at:
778 <http://www.cognisciences.com/accueil/outils/article/oura-lec-cp-outil-enseignant>.
- 779 Black, P. and Wiliam, D. (2009) ‘Developing the theory of formative assessment’, *Educational*
780 *Assessment, Evaluation and Accountability*(formerly: *Journal of Personnel Evaluation in Education*),
781 21(1), p. 5. Available at: <https://doi.org/10.1007/s11092-008-9068-5>.
- 782 Bogaerds-Hazenberg, S., Evers-Vermeul, J. and Bergh, H. van den (2019) ‘Teachers and researchers
783 as co-designers? A design-based research on reading comprehension instruction in primary
784 education’, *EDeR. Educational Design Research*, 3(1). Available at:
785 <https://doi.org/10.15460/eder.3.1.1399>.
- 786 Bondie, R.S., Dahnke, C. and Zusho, A. (2019) ‘How Does Changing “One-Size-Fits-All” to
787 Differentiated Instruction Affect Teaching?’, *Review of Research in Education*, 43(1), pp. 336–362.
788 Available at: <https://doi.org/10.3102/0091732X18821130>.
- 789 Bressoux, P. (2021) ‘A quelles conditions peut-on déployer à grande échelle les interventions qui
790 visent à améliorer les pratiques enseignantes ?’, in Galand, B. and Janosz, M., *Améliorer les*
791 *pratiques en éducation Qu’en dit la recherche ?* Presses universitaires de Louvain. Louvain-La-
792 Neuve.

- 793 Bricteux, S. and Quittre, V. (2021) *Résultats de PISA 2018 en Fédération Wallonie-Bruxelles Des*
794 *différences aux inégalités*. Service d'Analyse des Systèmes et des Pratiques d'enseignement.
- 795 Bryk, A.S. (2015) 'Accelerating How We Learn to Improve', *Educational Researcher*, 44(9), pp.
796 467–477. Available at: <https://doi.org/10.3102/0013189X15621543>.
- 797 Bryk, A.S. (2017) 'Accélérer la manière dont nous apprenons à améliorer', *Éducation et didactique*,
798 11(2), pp. 11–29. Available at: <https://doi.org/10.4000/educationdidactique.2796>.
- 799 Caena, F. (2011) *Teachers' continuing professional development*. European Commission, p. 21.
- 800 Calkins, L. (2017) *A Guide to the Reading Workshop: Primary Grades*. Teachers College Reading
801 and Writing Project. Columbia University.
- 802 Castro-Villarreal, F., Rodriguez, B.J. and Moore, S. (2014) 'Teachers' perceptions and attitudes
803 about Response to Intervention (RTI) in their schools: A qualitative analysis', *Teaching and Teacher*
804 *Education*, 40, pp. 104–112. Available at: <https://doi.org/10.1016/j.tate.2014.02.004>.
- 805 Cèbe, S. and Goigoux, R. (2018) 'Lutter contre les inégalités : outiller pour former les enseignants',
806 *Recherche & formation*, (87), pp. 77–96. Available at:
807 <https://doi.org/10.4000/rechercheformation.3510>.
- 808 Christ, T.J. *et al.* (2012) 'Curriculum-Based Measurement of Oral Reading: Quality of Progress
809 Monitoring Outcomes', *Exceptional Children*, 78(3), pp. 356–373. Available at:
810 <https://doi.org/10.1177/001440291207800306>.
- 811 Class, B. and Schneider, D. (2013) 'La Recherche Design en Education : vers une nouvelle
812 approche ?', *Frantice.net*, 7, p. 5.
- 813 Colognesi, S. and Gouin, J.-A. (2022) 'A typology of learner profiles to anticipate and guide
814 differentiation in primary classes', *Research Papers in Education*, 37(4), pp. 479–495. Available at:
815 <https://doi.org/10.1080/02671522.2020.1849376>.
- 816 Coppe, T. *et al.* (2018) 'Ouvrir la boîte noire du travail de préparation de l'enseignant : essai de
817 modélisation et d'illustration autour du choix et de l'évolution d'un document support de cours',
818 *Revue française de pédagogie. Recherches en éducation*, (204), pp. 17–31. Available at:
819 <https://doi.org/10.4000/rfp.8358>.
- 820 Cowan, C. and Maxwell, G. (2015) 'Educators' Perceptions of Response to Intervention
821 Implementation and Impact on Student Learning', *Journal of Instructional Pedagogies*, 16. Available
822 at: <https://eric.ed.gov/?id=EJ1069392> (Accessed: 8 July 2022).
- 823 Davis-Kean, P.E. (2005) 'The influence of parent education and family income on child
824 achievement: the indirect role of parental expectations and the home environment', *Journal of family*
825 *psychology: JFP: journal of the Division of Family Psychology of the American Psychological*
826 *Association (Division 43)*, 19(2), pp. 294–304. Available at: [https://doi.org/10.1037/0893-](https://doi.org/10.1037/0893-3200.19.2.294)
827 [3200.19.2.294](https://doi.org/10.1037/0893-3200.19.2.294).

- 828 Denessen, E. (2017) 'Dealing responsibly with differences. Socio-cultural backgrounds and
829 differentiation in education'. Universiteit Leiden. Available at:
830 <https://scholarlypublications.universiteitleiden.nl/access/item%3A2953124/view>.
- 831 Denessen, E. *et al.* (2022) 'Implicit measures of teachers' attitudes and stereotypes, and their effects
832 on teacher practice and student outcomes: A review', *Learning and Instruction*, 78, p. 101437.
833 Available at: <https://doi.org/10.1016/j.learninstruc.2020.101437>.
- 834 Deum, M. *et al.* (2007) 'Outil pour le diagnostic et la remédiation des difficultés d'acquisition de la
835 lecture en 1re et 2e années primaires'. Service général du Pilotage du système éducatif.
- 836 Deunk, M.I. *et al.* (2018) 'Effective differentiation Practices: A systematic review and meta-analysis
837 of studies on the cognitive effects of differentiation practices in primary education', *Educational
838 Research Review*, 24, pp. 31–54. Available at: <https://doi.org/10.1016/j.edurev.2018.02.002>.
- 839 Dietrichson, J. *et al.* (2017) 'Academic Interventions for Elementary and Middle School Students
840 With Low Socioeconomic Status: A Systematic Review and Meta-Analysis', *Review of Educational
841 Research*, 87(2), pp. 243–282. Available at: <https://doi.org/10.3102/0034654316687036>.
- 842 Dietrichson, J. *et al.* (2021) 'Targeted school-based interventions for improving reading and
843 mathematics for students with or at risk of academic difficulties in Grades K-6: A systematic review',
844 *Campbell Systematic Reviews*, 17(2). Available at: <https://doi.org/10.1002/cl2.1152>.
- 845 Donovan, M.S., Snow, C. and Daro, P. (2013) 'The SERP Approach to Problem-Solving Research,
846 Development, and Implementation', *Yearbook of the National Society for the Study of Education*,
847 112(2), pp. 400–425. Available at: <https://doi.org/10.1177/016146811311501411>.
- 848 Dupriez, V. (1999) 'La liberté pédagogique comme condition de la concurrence.', in Dumon, H. and
849 Collin, M., *Le décret du 24 juillet 1997 définissant les missions prioritaires de l'enseignement.
850 Approche interdisciplinaire*. Presses des FUSL. Bruxelles, Belgique, pp. 211–22. Available at:
851 <https://dial.uclouvain.be/pr/boreal/object/boreal:80478> (Accessed: 25 November 2022).
- 852 Eysink, T.H.S. and Schildkamp, K. (2021) 'A conceptual framework for Assessment-Informed
853 Differentiation (AID) in the classroom', *Educational Research*, 63(3), pp. 261–278. Available at:
854 <https://doi.org/10.1080/00131881.2021.1942118>.
- 855 Fédération Wallonie-Bruxelles. (2022). *Les indicateurs de l'enseignement*. Administration générale de
856 l'Enseignement. 17^e édition.
- 857 Filderman, M.J. *et al.* (2018) 'Data-Based Decision Making in Reading Interventions: A Synthesis
858 and Meta-Analysis of the Effects for Struggling Readers', *The Journal of Special Education*, 52(3),
859 pp. 174–187. Available at: <https://doi.org/10.1177/0022466918790001>.
- 860 Filderman, M.J. and Toste, J.R. (2018) 'Decisions, Decisions, Decisions: Using Data to Make
861 Instructional Decisions for Struggling Readers', *TEACHING Exceptional Children*, 50(3), pp. 130–
862 140. Available at: <https://doi.org/10.1177/0040059917740701>.
- 863 Fuchs, D. and Fuchs, L.S. (2017) 'Critique of the National Evaluation of Response to Intervention: A
864 Case for Simpler Frameworks', *Exceptional Children*, 83(3), pp. 255–268. Available at:
865 <https://doi.org/10.1177/0014402917693580>.

- 866 Gaitas, S. and Alves Martins, M. (2017) 'Teacher perceived difficulty in implementing differentiated
867 instructional strategies in primary school', *International Journal of Inclusive Education*, 21(5), pp.
868 544–556. Available at: <https://doi.org/10.1080/13603116.2016.1223180>.
- 869 van Geel, M. *et al.* (2019) 'Capturing the complexity of differentiated instruction', *School*
870 *Effectiveness and School Improvement*, 30(1), pp. 51–67. Available at:
871 <https://doi.org/10.1080/09243453.2018.1539013>.
- 872 Gennetian, L.A., Castells, N. and Morris, P.A. (2010) 'Meeting the basic needs of children: Does
873 income matter?', *Children and Youth Services Review*, 32(9), pp. 1138–1148. Available at:
874 <https://doi.org/10.1016/j.childyouth.2010.03.004>.
- 875 Gersten, R. *et al.* (2020) 'Meta-Analysis of the Impact of Reading Interventions for Students in the
876 Primary Grades', *Journal of Research on Educational Effectiveness*, 13(2), pp. 401–427. Available
877 at: <https://doi.org/10.1080/19345747.2019.1689591>.
- 878 Gersten, R., Jayanthi, M. and Dimino, J. (2017) 'Too Much, Too Soon? Unanswered Questions From
879 National Response to Intervention Evaluation', *Exceptional Children*, 83(3), pp. 244–254. Available
880 at: <https://doi.org/10.1177/0014402917692847>.
- 881 Godor, B.P. (2021) 'The Many Faces of Teacher Differentiation: Using Q Methodology to Explore
882 Teachers Preferences for Differentiated Instruction', *The Teacher Educator*, 56(1), pp. 43–60.
883 Available at: <https://doi.org/10.1080/08878730.2020.1785068>.
- 884 Goigoux, R. (2017) 'Associer chercheurs et praticiens à la conception d'outils didactiques ou de
885 dispositifs innovants pour améliorer l'enseignement', *Éducation et didactique*, 11(3), pp. 135–142.
886 Available at: <https://doi.org/10.4000/educationdidactique.2872>.
- 887 Goigoux, R., Renaud, J. and Roux-Baron, I. (2021) 'Comment influencer positivement les pratiques
888 pédagogiques de professeurs expérimentés ?', in Galand, B. and Janosz, M., *Améliorer les pratiques*
889 *en éducation : Qu'en dit la recherche ?*, pp. 67–76.
- 890 Gortazar, L., Martinez de Lafuente, D. and Vega-Bayo, A. (2022) 'Comparing teacher and external
891 assessments: Are boys, immigrants, and poorer students undergraded?', *Teaching and Teacher*
892 *Education*, 115, p. 103725. Available at: <https://doi.org/10.1016/j.tate.2022.103725>.
- 893 Gottheiner, D.M. and Siegel, M.A. (2012) 'Experienced Middle School Science Teachers'
894 Assessment Literacy: Investigating Knowledge of Students' Conceptions in Genetics and Ways to
895 Shape Instruction', *Journal of Science Teacher Education*, 23(5), pp. 531–557. Available at:
896 <https://doi.org/10.1007/s10972-012-9278-z>.
- 897 Greenfield, R. *et al.* (2010) 'Teachers' Perceptions of a Response to Intervention (RTI) Reform
898 Effort in an Urban Elementary School: A Consensual Qualitative Analysis', *Journal of Disability*
899 *Policy Studies*, 21(1), pp. 47–63. Available at: <https://doi.org/10.1177/1044207310365499>.
- 900 Hanin, V. *et al.* (2022) 'Association between prospective elementary school teachers' year of study
901 and their type of conception of intelligence', *International Journal of Educational Research*, 115, p.
902 102039. Available at: <https://doi.org/10.1016/j.ijer.2022.102039>.

- 903 Hanna, R. and Linden, L. (2009) 'Measuring Discrimination in Education'. National Bureau of
904 Economic Research (Working Paper Series). Available at: <https://doi.org/10.3386/w15057>.
- 905 Hargreaves, E. (2013) 'Inquiring into children's experiences of teacher feedback: reconceptualising
906 Assessment for Learning', *Oxford Review of Education*, 39(2), pp. 229–246. Available at:
907 <https://doi.org/10.1080/03054985.2013.787922>.
- 908 Hart, B. and Risley, T.R. (2003) 'The Early Catastrophe: The 30 million word gap by age 3',
909 *American educator*, 27(1), pp. 4–9.
- 910 Hebbeker, K. *et al.* (2022) 'Data-based decision-making in schools: Examining the process and
911 effects of teacher support.', *Journal of Educational Psychology*, 114(7), pp. 1695–1721. Available at:
912 <https://doi.org/10.1037/edu0000530>.
- 913 von Hippel, P.T. and Cañedo, A.P. (2022) 'Is Kindergarten Ability Group Placement Biased? New
914 Data, New Methods, New Answers', *American Educational Research Journal*, 59(4), pp. 820–857.
915 Available at: <https://doi.org/10.3102/00028312211061410>.
- 916 Hughes, C.A. and Dexter, D.D. (2011) 'Response to Intervention: A Research-Based Summary',
917 *Theory Into Practice*, 50(1), pp. 4–11. Available at: <https://doi.org/10.1080/00405841.2011.534909>.
- 918 Klute, M. *et al.* (2017) *Formative assessment and elementary school student academic achievement: A review of the evidence*. REL 2017-259. National Center For Education Evaluation and Regional
919 Assistance, p. 53. Available at: <http://ies.ed.gov/ncee/edlabs>.
- 921 Le Normand, M.-T., Parisse, C. and Cohen, H. (2008) 'Lexical diversity and productivity in French
922 preschoolers: Developmental, gender and sociocultural factors', *Clinical linguistics & phonetics*, 22,
923 pp. 47–58. Available at: <https://doi.org/10.1080/02699200701669945>.
- 924 Lemons, C.J., Kearns, D.M. and Davidson, K.A. (2014) 'Data-Based Individualization in Reading:
925 Intensifying Interventions for Students With Significant Reading Disabilities', *TEACHING*
926 *Exceptional Children*, 46(4), pp. 20–29. Available at: <https://doi.org/10.1177/0040059914522978>.
- 927 Magnuson, K. and Shager, H. (2010) 'Early education: Progress and promise for children from low-
928 income families', *Children and Youth Services Review*, 32(9), pp. 1186–1198. Available at:
929 <https://doi.org/10.1016/j.childyouth.2010.03.006>.
- 930 Morrisette, J. (2011a) 'Formative Assessment: Revisiting the territory from the point of view of
931 teachers', *McGill Journal of Education*, 46(2), pp. 247–265. Available at:
932 <https://doi.org/10.7202/1006438ar>.
- 933 Morrisette, J. (2011b) 'Vers un cadre d'analyse interactionniste des pratiques professionnelles',
934 *Recherches qualitatives*, 30(1), pp. 10–32. Available at: <https://doi.org/10.7202/1085478ar>.
- 935 Morrisette, J. and Guignon, S. (2016) 'Trois zones de coconstruction de savoirs professionnels
936 issues des médiations de débats en groupe', *Communiquer. Revue de communication sociale et*
937 *publique*, (18), pp. 117–130. Available at: <https://doi.org/10.4000/communiquer.2085>.

- 938 Nadeau, A. (2021) ‘Conceptions d’enseignants du primaire sur leur rôle de passeur culturel : effets de
939 dispositifs d’intégration de la dimension culturelle à l’école québécoise’, *Recherches qualitatives*,
940 40(1), pp. 128–153. Available at: <https://doi.org/10.7202/1076350ar>.
- 941 Neitzel, A.J. *et al.* (2021) ‘A Synthesis of Quantitative Research on Programs for Struggling Readers
942 in Elementary Schools’, *Reading Research Quarterly*, 57(1), pp. 149–179. Available at:
943 <https://doi.org/10.1002/rrq.379>.
- 944 Nisbett, R.E. *et al.* (2012) ‘Intelligence: New findings and theoretical developments.’, *American*
945 *Psychologist*, 67(2), pp. 130–159. Available at: <https://doi.org/10.1037/a0026699>.
- 946 Oslund, E.L., Elleman, A.M. and Wallace, K. (2021) ‘Factors Related to Data-Based Decision-
947 Making: Examining Experience, Professional Development, and the Mediating Effect of Confidence
948 on Teacher Graph Literacy’, *Journal of Learning Disabilities*, 54(4), pp. 243–255. Available at:
949 <https://doi.org/10.1177/0022219420972187>.
- 950 Peters, M.T., Hebbecke, K. and Souvignier, E. (2022) ‘Effects of Providing Teachers With Tools for
951 Implementing Assessment-Based Differentiated Reading Instruction in Second Grade’, *Assessment*
952 *for Effective Intervention*, 47(3), pp. 157–169. Available at:
953 <https://doi.org/10.1177/15345084211014926>.
- 954 Prenger, R. and Schildkamp, K. (2018) ‘Data-based decision making for teacher and student
955 learning: a psychological perspective on the role of the teacher’, *Educational Psychology*, 38(6), pp.
956 734–752. Available at: <https://doi.org/10.1080/01443410.2018.1426834>.
- 957 Puzio, K., Colby, G.T. and Algeo-Nichols, D. (2020) ‘Differentiated Literacy Instruction:
958 Boondoggle or Best Practice?’, *Review of Educational Research*, 90(4), pp. 459–498. Available at:
959 <https://doi.org/10.3102/0034654320933536>.
- 960 Quinn, D.M. and Kim, J.S. (2017) ‘Scaffolding Fidelity and Adaptation in Educational Program
961 Implementation: Experimental Evidence From a Literacy Intervention’, *American Educational*
962 *Research Journal*, 54(6), pp. 1187–1220. Available at: <https://doi.org/10.3102/0002831217717692>.
- 963 Quittre, V., Dupont, V. and Lafontaine, D. (2021) ‘Des enseignants parlent aux enseignants Résultats
964 de l’enquête TALIS 2018’. Service d’Analyse des Systèmes et des Pratiques d’enseignement.
- 965 Renard, F. *et al.* (2022) ‘De la structure légère de pilotage au Pacte pour un Enseignement
966 d’excellence. Le glissement progressif d’un pilotage incitatif à un pilotage par les résultats et la
967 reddition de comptes en Belgique francophone’, *Les Dossiers des Sciences de L Éducation*, 45, pp.
968 33–56.
- 969 Renaud, J. (2020) ‘Évaluer l’utilisabilité, l’utilité et l’acceptabilité d’un outil didactique au cours du
970 processus de conception continuée dans l’usage’, *Éducation et didactique*, (14–2), pp. 65–84.
971 Available at: <https://doi.org/10.4000/educationdidactique.6756>.
- 972 Roy, A., Guay, F. and Valois, P. (2013) ‘Teaching to address diverse learning needs: development
973 and validation of a Differentiated Instruction Scale’, *International Journal of Inclusive Education*,
974 17(11), pp. 1186–1204. Available at: <https://doi.org/10.1080/13603116.2012.743604>.

- 975 Scarborough, H.S. (2005) 'Developmental Relationships between Language and Reading:
976 Reconciling a Beautiful Hypothesis with Some Ugly Fact', in Catts, H. W., *The connections between*
977 *language and reading disabilities*. Edited by A. G. Kamhi. Lawrence Erlbaum Associates Publishers,
978 pp. 3–24.
- 979 Schelling, N. and Rubenstein, L.D. (2021) 'Elementary teachers' perceptions of data-driven decision-
980 making', *Educational Assessment, Evaluation and Accountability*, 33(2), pp. 317–344. Available at:
981 <https://doi.org/10.1007/s11092-021-09356-w>.
- 982 Schildkamp, K. (2019) 'Data-based decision-making for school improvement: Research insights and
983 gaps', *Educational Research*, 61(3), pp. 257–273. Available at:
984 <https://doi.org/10.1080/00131881.2019.1625716>.
- 985 Schillings, P. *et al.* (2017) *PIRLS 2016 : Note de synthèse*, p. 22. Available at:
986 <http://hdl.handle.net/2268/216693>.
- 987 Slates, S.L. *et al.* (2012) 'Counteracting Summer Slide: Social Capital Resources Within
988 Socioeconomically Disadvantaged Families', *Journal of Education for Students Placed at Risk*
989 (*JESPAR*), 17(3), pp. 165–185. Available at: <https://doi.org/10.1080/10824669.2012.688171>.
- 990 Slavin, R.E. *et al.* (2011) 'Effective programs for struggling readers: A best-evidence synthesis',
991 *Educational Research Review*, 6(1), pp. 1–26. Available at:
992 <https://doi.org/10.1016/j.edurev.2010.07.002>.
- 993 Snow, C.E. (2015) '2014 Wallace Foundation Distinguished Lecture: Rigor and Realism: Doing
994 Educational Science in the Real World', *Educational Researcher*, 44(9), pp. 460–466. Available at:
995 <https://doi.org/10.3102/0013189X15619166>.
- 996 Sprietsma, M. (2013) 'Discrimination in grading: experimental evidence from primary school
997 teachers', *Empirical Economics*, 45(1), pp. 523–538. Available at: [https://doi.org/10.1007/s00181-](https://doi.org/10.1007/s00181-012-0609-x)
998 [012-0609-x](https://doi.org/10.1007/s00181-012-0609-x).
- 999 Stecker, P.M., Lembke, E.S. and Foegen, A. (2008) 'Using Progress-Monitoring Data to Improve
1000 Instructional Decision Making', *Preventing School Failure: Alternative Education for Children and*
1001 *Youth*, 52(2), pp. 48–58. Available at: <https://doi.org/10.3200/PSFL.52.2.48-58>.
- 1002 Taylor, B. *et al.* (2022) 'Attainment grouping in English secondary schools: A national survey of
1003 current practices', *Research Papers in Education*, 37(2), pp. 199–220. Available at:
1004 <https://doi.org/10.1080/02671522.2020.1836517>.
- 1005 The Design-Based Research Collective (2003) 'Design-Based Research: An Emerging Paradigm for
1006 Educational Inquiry', *Educational Researcher*, 32(1), pp. 5–8. Available at:
1007 <https://doi.org/10.3102/0013189X032001005>.
- 1008 UNICEF Office of Research (2016) *Fairness for children: a league table of inequality in child well-*
1009 *being in rich countries*. Florence: UNICEF Office of Research - Innocenti (Innocenti Report Card,
1010 13).

- 1011 Van der Kleij, F.M. *et al.* (2015) ‘Integrating data-based decision making, Assessment for Learning
1012 and diagnostic testing in formative assessment’, *Assessment in Education: Principles, Policy &
1013 Practice*, 22(3), pp. 324–343. Available at: <https://doi.org/10.1080/0969594X.2014.999024>.
- 1014 Van Nieuwenhoven, C. and Colognesi, S. (2015) ‘Une recherche collaborative sur
1015 l’accompagnement des futurs instituteurs : un levier de développement professionnel pour les maîtres
1016 de stage’, *e-Jiref*, 1, pp. 103–121.
- 1017 Van Norman, E.R., Nelson, P.M. and Parker, D.C. (2018) ‘A comparison of nonsense-word fluency
1018 and curriculum-based measurement of reading to measure response to phonics instruction.’, *School
1019 Psychology Quarterly*, 33(4), pp. 573–581. Available at: <https://doi.org/10.1037/spq0000237>.
- 1020 Vijfeijken, M.V. *et al.* (2021) ‘Equity, Equality, and Need: A Qualitative Study into Teachers’
1021 Professional Trade-Offs in Justifying Their Differentiation Practice’, *Open Journal of Social
1022 Sciences*, 9(8), pp. 236–257. Available at: <https://doi.org/10.4236/jss.2021.98017>.
- 1023 Visscher, A.J. (2021) ‘On the value of data-based decision making in education: The evidence from
1024 six intervention studies’, *Studies in Educational Evaluation*, 69, p. 100899. Available at:
1025 <https://doi.org/10.1016/j.stueduc.2020.100899>.
- 1026 Wang, S., Rubie-Davies, C.M. and Meissel, K. (2018) ‘A systematic review of the teacher
1027 expectation literature over the past 30 years’, *Educational Research and Evaluation*, 24(3–5), pp.
1028 124–179. Available at: <https://doi.org/10.1080/13803611.2018.1548798>.
- 1029 Wayman, J.C., Jimerson, J.B. and Cho, V. (2012) ‘Organizational considerations in establishing the
1030 Data-Informed District’, *School Effectiveness and School Improvement*, 23(2), pp. 159–178.
1031 Available at: <https://doi.org/10.1080/09243453.2011.652124>.
- 1032 Yang, C. *et al.* (2021) ‘Testing (quizzing) boosts classroom learning: A systematic and meta-analytic
1033 review’, *Psychological Bulletin*, 147(4), pp. 399–435. Available at:
1034 <https://doi.org/10.1037/bul0000309>.
- 1035 Yin, Y., Tomita, M.K. and Shavelson, R.J. (2014) ‘Using Formal Embedded Formative Assessments
1036 Aligned with a Short-Term Learning Progression to Promote Conceptual Change and Achievement
1037 in Science’, *International Journal of Science Education*, 36(4), pp. 531–552. Available at:
1038 <https://doi.org/10.1080/09500693.2013.787556>.
- 1039 Zorman, M. *et al.* (2015) ‘« PARLER » : un dispositif pour prévenir les difficultés scolaires’, *Revue
1040 française de pédagogie*, (193), pp. 57–76. Available at: <https://doi.org/10.4000/rfp.4890>.
- 1041

6 Manuscript Formatting

6.1 Tables

Table 1. Overview of the participants and their main characteristics

Given name	Role	Gender	Years of experience	Other qualifications and additional experience	SEI level of the school ¹
Sophia	Teacher 1st grade	F	8	Master in educational sciences (In progress)	2 (Strongly disadvantaged)
Victoria	Teacher 1st and 2 nd grade	F	24	Higher education degree in pedagogy Co-author of a teacher manual for reading/writing	19 (Strongly advantaged)
Carol	Teacher 1st and 2 nd grade	F	18	Master in educational sciences Co-author of a teacher manual for reading/writing	4 (Strongly disadvantaged)
Lucy	Teacher 1st and 2 nd grade	F	9	Master in educational sciences (In progress) Participated in a study on the effect of co-teaching on reading performance	8 (Disadvantaged)
George	Teacher 2 nd grade	M	7	Master in educational sciences	1 (Strongly disadvantaged)
Mary	PhD student – Former teacher 1 st grade	F	/		
Eve	Researcher	F	/		

¹ In the French-speaking part of Belgium, each school is ranked on a scale from one to twenty depending on the socio-economical background of the pupils (Socio-Economic Index - SEI). It is computed based among other things on the incomes and educational level of pupils' parents (Fédération Wallonie-Bruxelles, 2022).

1049 Table 2. Summary of the focus groups

Focus Group	Presence							Length	Modality	Main objectives
	Sophia	Victoria	Carol	Laura	George	Mary	Eve			
1	X	X	X	X	X	X	X	3h	In-person	Getting to know each other Brief theoretical explanation Exploration of the existing tools
2	X		X	X	X		X	3h	In-person	Development of the first prototype (individual rubric)
3	X	X	X	X	X		X	1h30	Online	Improvement of the individual rubric Development of the whole-class tools
4		X	X	X	X		X	3h	Online	Finetuning the prototypes Writing the appendix for future users Views on the collaboration process

1050

1051

1052 Table 3. Overview of results crossing participant needs with the interactionist framework

Need	Area	Details
Perceived usefulness	Shared	Identifying pupils' level differences
		Differentiate
		Log information
	Accepted	Summarize essential steps for learning to read Different tools serving different goals
Limiting the workload	Shared	Progress monitoring is time-consuming
		Tools' format facilitates easy use
Balancing workload and perceived usefulness	Disputed	Two groups in implementing the tool for an individual evaluation
Flexibility	Shared	Easily adjustable by teachers
		Teacher expertise valued
		Adjustable to pupils' level and needs
	Accepted	Liberty in the format of the evaluation

1053

1054

Supplementary Material

Co-creating tools to monitor first graders' progress in reading: A balancing act between perceived usefulness, flexibility and workload

Eve Francotte^{1*}, Stéphane Colognesi^{1†}, Liesje Coertjens^{1†}

¹Psychological Sciences Research Institute, UCLouvain, Louvain-la-Neuve, Belgium

†: Equal contribution and last authorship

*** Correspondence:**

Eve Francotte (she/her)

eve.francotte@uclouvain.be

1 Supplementary Data

2 Supplementary Figures and Tables

The tools presented here are translated from French.

Name:

Rubrics for reading component

Class:

		.../...	.../...	.../...	.../...	.../...
Phonological awareness:						
Recognize a sound in different spoken words						
Locate the position of a phoneme or a syllable						
Match words that rhyme						
Identify the first letter or first syllable of a word						
Identify syllables of a word						
Identify phonemes of a word						
Rapid automatized naming ¹ :						
Name quickly a series of colors (or pictures, objects)						
Respect the reading direction, from left to right and from top to bottom						
Decode to read :						
Letter-sound correspondence	Vowels					
	Long consonants					
	Short consonants					
Letters: a – e – i – o – u – y – é – è – ê / l – r – m – s – f – n – v – j – z – ch / p – t – c – k – q(u) – d – b – g – x – w – h						
Recognize letters in different shapes (cursive, capital, ...)	Vowels					
	Long consonants					
	Short consonants					

Being able to merge	Consonant + Vowel					
	Vowel + Consonant					
	Double cons. + Vowel					
Decode words with	Simple sounds					
	Complex sounds					
Sounds : ou – on, om – en, an, em, am – oi – ui – au, eau – in, ain, aim, ein – ien – ei, ai – e + 2C – er, ez – oin – eu, eur – gn – ph – tion – ille – ouil – ail – eil – euil						
Global recognition of function words						
Decode to understand :						
Decode and understand words with	Simple sounds ¹					
	Complex sounds ¹					
Understand sentences	Short, without complex sounds					
	Short with complex sounds					
	Long without inference					
Understand a short text independently						

¹See Appendix for further details

Supplementary Figure 1. Criterion-based rubric for reading components.

Progress monitoring in reading

Class:

[illegible]

Id. = Identify / *See appendix

[illegible]

Supplementary Figure 2. Whole-class tool for reading components.

Progress monitoring in reading

Class:

[illegible]

[illegible]

[illegible]

Supplementary Figure 3. Whole-class tool for letter-sound correspondence.

Progress monitoring in reading

Class:

[illegible]

Supplementary Figure 4. Blank version.

Appendix from the tools

Appendix: Progress monitoring in reading

The purpose of these tools is to help teachers to monitor their first-grade pupils' progress in reading. They were created by a group of teachers as part of a research project at the UCLouvain (Louvain-la-Neuve, French-speaking part of Belgium).

These tools are intended to be as flexible as possible in order to match all teaching practices. This is why there is no precise guide on how to complete each tool. Moreover, you are free (and encouraged) to modify criteria, layout, list of sounds This appendix provides some necessary complements to properly use the tool. These tools do not intend to be exhaustive.

The **rubric on reading components** aims to monitor progress of one particular pupil. Each column represents an assessment date. The assessor is asked to indicate the level of acquisition of criterion (e.g., Not Invested/Not acquired/Acquired/in the process of being acquired).

The two **whole-class tools** (i.e., whole-class tool for reading component and the whole-class tool for letter-sound correspondence) aim to monitor progress at the class-level: for each pupil the assessor indicates whether the criterion (e.g., the correspondence between a letter and a sound) is acquired. While these tools are less precise for an individual pupil, they can be useful to create ability groups.

The **blank version** allows you to define the list of letter-sound correspondences yourself, in line with your preference, and the order in which these are correspondences are tackled in your class.

Here are some further details on the rubric on reading components:

Rapid automatized naming is an under-researched predictor of reading skills. It refers to "the time required for a child to quickly and accurately name an array of well-known visual stimuli (usually letters, digits, objects, or colors)" (Araújo et al., 2015, p.868). An example of an exercise is presented below (Beaudry, 2010, figure 1): pupils must name the

drawings as quickly as possible, from left to right (grapes – hat – scissors -...). Please note that it is important that the pupil knows the vocabulary of all the drawing, otherwise the evaluation of this criterion risks to be biased.

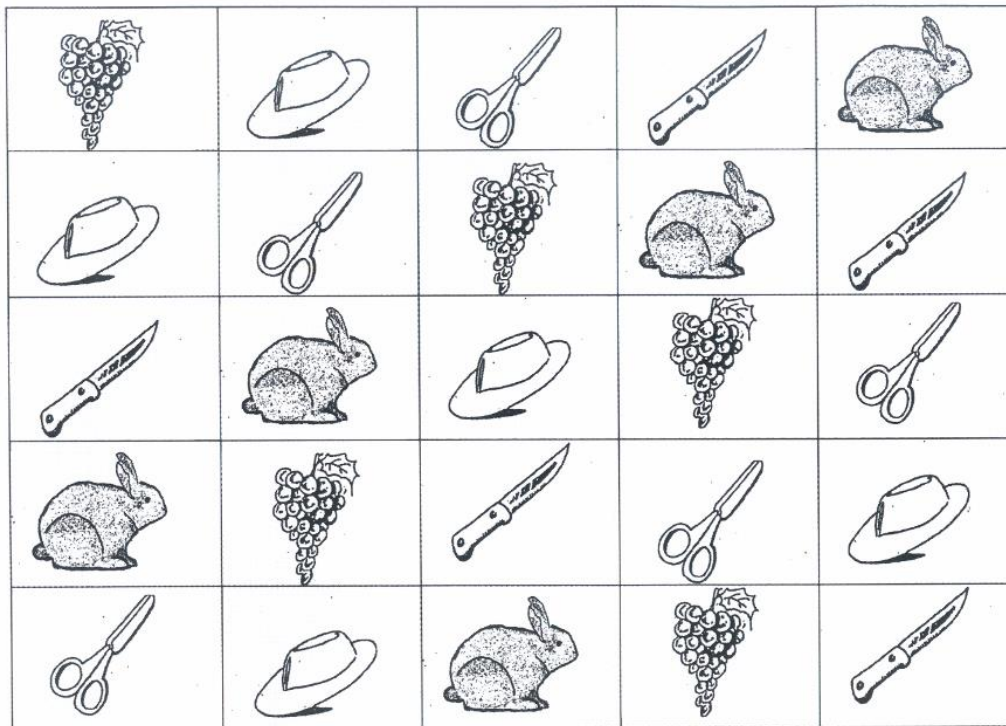


Figure 1. Example of a Rapid Automatized Naming exercise (Reprinted from Beaudry, 2010, p.70)

In the rubrics for reading component, it is recommended to circle the letter/sound when the letter-sound correspondence is mastered. In rubrics, "simple sound" refers to a phoneme linked to one letter (a – l – b – l - ...), while «complex sound» refers to a phoneme linked to several letters (oi – oo - ...). « Function words » in the rubric refers to the list created by Giasson (2003) (Figure 2). This list of French function words is presented below.

à, au, aux	et	où	son, sa, ses
aller **	être**	par	sur
autre	eux	pas	t', te, tu, toi
avec	faire **	plus	ton, ta, tes
avoir **	il, ils	pour	tout, toute, tous
bien	j', je, m', me, moi	pouvoir **	un, une, uns, unes
c', ce, cet, cette, ces	l', le, la, les	prendre **	venir **
comme	leur, leurs	qu', que	voir **
d', de, du, des	lui	qui	votre, vos
dans	mais	s', se	vouloir **
dire **	mon, ma, mes	sans	vous
donner **	n', ne	savoir **	y
elle, elles	notre, nos	si	
en	on	soi	

** Y compris les formes conjuguées du verbe mentionnées dans les tableaux de savoirs et de savoir-faire. (Giasson, 2003)

Figure 2. List of function words (Reprinted from Giasson, 2003)

References:

- Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, 107(3), 868-883. <https://doi.org/10.1037/edu0000006>
- Beaudry, C. (2010). *Les dyslexies, troubles du langage écrit*. <https://slideplayer.fr/slide/4837885/>
- Giasson, J. (2003) *La lecture : de la théorie à la pratique*. 2^e édition. Boucherville : Gaëtan Morin.