Effects of a formative assessment system on early reading development

Stefan Gustafson, Thomas Nordström, Ulrika B Andersson, Linda Fälth and Ingvar Martin

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-161353

N.B.: When citing this work, cite the original publication.

Original publication available at:
Copyright: Project Innovation
http://www.projectinnovation.biz/education_2006.html
Effects of a formative assessment system on early reading development

Stefan Gustafson
Linköpings University

Thomas Nordström
Linnaeus University

Ulrika B Andersson
Linköping University

Linda Fälth
Linnaeus University

Martin Ingvar
Karolinska institutet

Short title: Formative assessment and reading
Abstract

We present quantitative results from the pilot-year of a large scale Swedish educational project in reading development called LegiLexi, inspired by research within the Response to intervention and Formative assessment traditions. The vision of the project is that every pupil should reach adequate reading skills at the end of grade 3 in primary school. LegiLexi contains a formative assessment tool and a teacher course, which are linked together. We describe LegiLexi and analyze quantitative effects of the pilot year regarding reading development for pupils in grade 1. The design included three conditions; full access to LegiLexi, access only to the formative assessment tool, and control.

Results showed that the group with full access to LegiLexi improved their word decoding and reading comprehension the most. For language comprehension, the Formative assessment only group showed the highest improvements. Thus, the features of LegiLexi seem to help enhance critical reading skills. Some changes will be made in the project to strengthen methodological aspects and further facilitate pupils’ reading development.

Keywords: formative assessment, intervention, primary school, reading instruction
Effects of a formative assessment system on early reading development

During the last decades, research on reading development and on effective reading instruction has made significant progress. There is now massive evidence on the critical role of phonological awareness, decoding (i.e., building links between letters and sounds), and language comprehension in developing reading comprehension (Ehri et al., 2001; Fletcher, Lyon, Fuchs, & Barnes, 2007; Frost, Madsbjerg, Niedersøe, Olofsson, & Sørensen, 2005; Anonymized authors, 2011; Torgesen, Alexander, Wagner, Rashotte, Voeller, & Conway, 2001). More specifically, studies have shown that early reading development heavily depends on phonological awareness and decoding and that language comprehension plays an increasingly important role for reading comprehension (de Jong & van der Leij, 2002; Anonymized authors, 2015).

It is also clear that reading instruction should be dynamic rather than static and should change depending on what the child already knows and needs to learn to advance further (Connor et al. 2009; Grigorenko, 2009; Vaughn & Fuchs, 2003). For example, a child who does not even know the letters of the alphabet would need to acquire that knowledge in order to decode words. A child who has already mastered automatic decoding of most words would still need to learn advanced reading comprehension by expanding vocabulary, improve the ability to make inferences and understand pragmatic aspects of written language. This developmental perspective has been implemented in the formative assessment tool and teacher course in LegiLexi, by using repeated measurements of different reading skills and suggesting reading instruction based on the test results.

Taken together, the two aspects regarding the content and form of instruction, respectively, provide a general framework for how to develop a successful and individualized educational system for reading instruction. This developmental and dynamic view is in line with research within Response to intervention (RTI) tradition.
as well as research within the formative assessment (FA) tradition. RTI is an educational model used mainly in the USA (Fuchs, Fuchs, & Compton, 2012; Mellard, McKnight, & Jordan, 2010). The idea is that no child should be left behind in their reading development. One fundamental principle of RTI is that the response of the individual child can be seen as an assessment of the particular intervention used for that child (Fuchs et al., 2012; Grigorenko, 2009). If the response is low, some changes in the instruction will be made, either by changing the content or form (organization) of instruction. Regarding content, it is argued that only evidence based methods should be used (accompanied by assessments of responses). Regarding form, different tiers of instruction is proposed, from general instruction in the classroom in Tier 1, to focused instruction in smaller groups in Tier 2, to more individualized and intense instruction in Tier 3 (Anonymized authors, 2014; Mellard et al., 2010).

Formative assessment also highlights the forward-looking aspect of assessment but researchers here focus more on feedback to students, which is regarded as a means to bridge the gap between the pupil’s current level and a goal (Black & William, 1998, 2018; Sadler, 1989; William, 2011). Thus, in the formative assessment tradition, more focus is on students’ self-regulating abilities in response to feedback from a teacher than on teacher’s decisions regarding learning opportunities. In relation to reading instruction, RTI might provide a more appropriate framework for skills that are relatively easy to measure quantitatively by means of for example letter knowledge tests or decoding tests. Conversely, FA might have more to offer in more complex tasks, such as elaborating on the meanings of a text, where self-regulation after feedback and interaction with the teacher and peers can be expected to facilitate learning.

In the present study, we will present quantitative results from the pilot-year of LegiLexi (https://www.legilexi.org). The project is run pro-bono and the developers and research team were inspired by research both from the fields of RTI and FA. The vision of LegiLexi is that all children in Swedish grade 1-3 should develop adequate reading skills at the end of grade 3. In order to achieve this, LegiLexi includes assessments which provide information about how successful past educational efforts
have been (i.e., response to intervention related to learning goals) and are used to suggest what should be focused next in reading instruction for each participating pupil. The project is organized by a foundation and the educational content is available for free for participating schools. LegiLexi involves participation with researchers in the field of reading development and instruction and also cooperation with a large number of school leaders and teachers. The complete version of LegiLexi will consist of three parts; a formative assessment tool, a course for teachers and an information library. The formative assessment tool is linked to the course material. Teachers get feedback on the progress of individual children and progress at the classroom level. Teachers also receive suggestions regarding reading instruction for particular children based on the profile of their reading skills. These suggestions are based on learning objectives found in the Swedish curriculum for grade 1 and 3 (Skolverket, 2011). Thus, recommendations are not based on relative comparisons among children but on learning goals. The course material contains chapters related to the recommendations. For example, there is one chapter about decoding and in that chapter there are examples of lessons targeting decoding skills (see Materials section for a more detailed description). The third part of LegiLexi, the information library, is still in development and will contain articles and other material related to reading development and reading instruction.

At the time of the present study, the formative assessment tool was already implemented, the teacher course was partly implemented and the information library was not yet implemented. We will provide quantitative results from the pilot year regarding reading development for pupils in grade 1. In an attempt to separate effects from different parts of LegiLexi we present results for two different experimental conditions; full access to LegiLexi (both the formative assessment tool and the teacher course) and access only to the formative assessment tool. Results will also be compared to a control condition (regular classroom instruction which was monitored but not interfered with).
Aims
The overall aim is to assess the effects of participating in the pilot year of LegiLexi on reading development in grade 1. We will analyze if both experimental groups show larger improvements than controls on measures of decoding, language comprehension and reading comprehension during grade 1 and also compare the two experimental conditions. Results from three test occasions, at pretest, midtest and posttest in grade 1, will be presented and analyzed. A secondary aim is to contribute with knowledge regarding the implementation of an educational project such as LegiLexi and discuss the results in relation to theories and concepts within RTI and FA.

Although the teacher course included theoretical concepts and suggestions for evidence based reading instruction, the results primarily reflect the use of a formative educational model rather than the use of any specific method for reading instruction (i.e., form rather than content). Teachers received different suggestions of reading instruction for different pupils depending on the pattern of test results and we analyze the effects of this attempt to help individualize reading instruction. The study contributes with knowledge about how assessments in the form of reading tests and associated recommendations and course material for teachers can guide reading instruction and facilitate early reading development.

Method
Participants
Participants were randomly assigned to groups on the school level. The main reason for this was to avoid dissemination between groups who had varied access to the LegiLexi material (full access, FA only, or no access). Randomization on the child level was not possible since the formative assessment tool and course material was directed at teachers. We assigned more children to the full LegiLexi group compared to FA only, since this was a pilot study and we wanted to receive comprehensive feedback on both the formative assessment tool and the teacher course. A total of 21 schools/teachers and 511 children participated in all three test occasions. In the full LegiLexi group there
were 8 schools/teachers and 217 children. In the LegiLexi – FA only group there were 4 schools/teachers and 86 children. In the control condition there were 9 schools/teachers and 208 children.

Test materials and test procedure
As part of LegiLexi, teachers administered the testing. They followed a comprehensive written test manual (Anonymized authors, 2017). Each assessment session was divided into two parts, group tests (pre-reading skills, language comprehension and reading comprehension tests) and second, the two word decoding tests which were administered individually (see below). Assessments were carried out three times (T1 in August, T2 in December and T3 in May) during grade 1 in primary school and all tests were included on each test occasion. Most tests were developed by the research team specifically for LegiLexi and therefore only test-retest correlations which were obtained in the present study are provided below. The word reading test was based on a Swedish test called H4. All tests were in Swedish and specifically adapted to the age of participants.

Phonological awareness. The test consisted of three sections. For the first four items, the teacher presented initial phonemes of words and the child was asked to select a picture among five depicting a word that begins with that sound. For the next four items the procedure was the same but instead directed at final phonemes. For the last eight items, the teacher said the individual phonemes of words, such as ‘c-a-t’, starting with three-letter words with successively longer words (the longest word had seven phonemes). The child was asked to choose the picture that formed the word for each item. The total number of correctly solved items was used for assessing phonological awareness with a maximum score of 16 points. Since data was not normally distributed we used a nonparametric test to assess reliability. Test-retest correlation between T1 and T2 was Spearman’s \( \rho = .44, p < .01 \).
**Letter-sound knowledge.** The teacher presented sounds that correspond to one of ten letters, e.g. ‘lll’ as in ‘lamp’. Uppercase letters were used for four items, lowercase letters for six items and two items had both an initial uppercase letter and lowercase letters. The number of correctly identified letters was used as a measure of letter-sound knowledge with a maximum of 12 points. Since data was not normally distributed we used a nonparametric test to assess reliability. Test-retest correlation between T1 and T2 was Spearman’s rho=.49, p<.01.

**Word reading.** The student was instructed to read as many words as possible from a list containing words with increasing length (from 2 to 6 letters) and difficulty within one minute. The test was performed individually and the test leader marked correct/incorrect answers on a score sheet containing the word list. The number of correctly read words was used as a measure of word reading and there was no maximum score since it was impossible to read all words in the list in one minute. Test-retest correlation between T1 and T2 was r=.91, p<.01.

**Pseudo word reading.** The test had the same format as the word reading test above but included pseudo words instead of real words. The number of correctly read pseudo words was used as a measure of pseudo word reading and there was no maximum score. Test-retest correlation between T1 and T2 was r=.82, p<.01.

**Listening comprehension – words.** The teacher said a word (both nouns and verbs were used) and the student were to identify the picture that best corresponded to the word among five pictures. Words were successively more difficult. There were 24 items and thus maximum score was 24 points. Test-retest correlation between T1 and T2 was r=.63, p<.01.

**Listening comprehension – sentences.** The teacher read sentences aloud and children were asked to select the picture that best corresponded to the sentence among five alternatives. Sentences were successively longer and more complicated. Maximum score was 12 points. Test-retest correlation between T1 and T2 was r=.56, p<.01.
Reading comprehension – pictures. Participants read sentences and answered 12 questions about the content by choosing between five pictures. The sentences become progressively longer and more difficult. Time constraint was 5 minutes. Maximum score was 12 points. Test-retest correlation between T1 and T2 was \( r = .69, p < .01 \).

Reading comprehension – text. The children read four texts, with three questions attached to each, and choose the answer that best corresponded to the content and meaning of the text. Time constraint was 7 minutes and maximum score was 18 points. Test-retest correlation between T1 and T2 was \( r = .76, p < .01 \).

LegiLexi

LegiLexi\(^1\) (https://www.legilexi.org/) was developed as an attempt to stop the decline in reading ability among Swedish pupils as shown in international comparisons (PISA-reports 2006; 2009 and 2012). The overall purpose is to help all children develop adequate reading skills during grade 1-3. The present study reports results from the pilot year (autumn 2015-spring 2016). During the pilot year, teachers had access to two features of LegiLexi, the formative assessment tool and the teacher course with corresponding exercises.

Formative assessment tool: Results from the 8 tests described above were used in the formative assessment tool. Teachers received computer generated compilations of those scores. To simplify interpretations, scores on individual tests were grouped together into three broad variables; pre-reading skills and decoding (phonological awareness, letter-sound knowledge, word reading and pseudo word reading), language comprehension (listening comprehension words and listening comprehension sentences, and reading comprehension (reading comprehension pictures and reading comprehension text).

---

\(^1\) LegiLexi is a private pro-bono initiative with the aim to supply teachers with the responsibility to teach reading with a practical and evidence-based formative assessment tool that is free of charge.
Teachers received feedback on the performance of individual children as well as aggregated information on the classroom level. The information for individual children had the form of graphs showing scores on the three variables described above. The information on the classroom level had the form of a class list showing each pupil's skill level on decoding, comprehension, and reading comprehension, as well as a matrix showing all pupils of the class grouped according to skill level on pre-reading skills, decoding, vocabulary, comprehension, and reading comprehension. Teachers also received information regarding the gap between current level and the long-term goal, which was set to reflect the curriculum's goals for reading skills in grade 3. After test session two and three teachers also received information regarding change between test sessions in the form of graphs for individual children.

Test results were meant to be used as a formative tool to guide reading instruction and teachers also received suggestions for what their instruction should focus on for individual children after the second test occasion (T2). Suggestions were based on empirical evidence regarding the relative influence of pre-reading skills, decoding, language comprehension, and reading comprehension during development. If a child had not yet mastered basic pre-reading skills, recommendations were directed towards those skills (phonological awareness or letter-sound knowledge). If those skills were interpreted as sufficient, suggestions were aimed at one of the two components in the Simple view of reading, decoding, or language comprehension. If decoding and language comprehension were above a predetermined threshold, suggestions were aimed at advanced reading comprehension instruction. Suggestions were automatically generated by an algorithm and were based on predetermined thresholds. Those were discussed among the research team and project leaders and were intended to reflect the objectives stated in the Swedish curriculum. Participating teachers also provided information about the perceived validity of the recommendations during the pilot year. Suggestions were brief in nature. As an example, the formative tool provided this suggestion generated for a pupil who showed strengths in language comprehension skills relative to her word decoding ability:
Focus area for the student: alphabetic decoding.

Individual recommendation: The student has a well-developed understanding of language, but needs to develop the decoding ability. Focus on developing the student's word decoding ability in order to get the student started with his/her own reading.

Since the participating children were in an early stage of their reading development but already had gained some letter-sound knowledge, the most common recommendation was alphabetic decoding (for LegiLexi FA 74% and for Full LegiLexi 83%).

Importantly, teachers who belonged to the full access to LegiLexi group could find useful information about decoding and decoding instruction, as well as more general information about reading and reading instruction, in the course material (see below).

Teacher course: A comprehensive course material in the form of 12 chapters was included in LegiLexi. Chapters were written by Swedish researchers and educators in the field of reading and reading instruction. Teachers in the LegiLexi group had access to and were encouraged to inform themselves about a variety of topics in the chapters, such as What is reading?, Decoding, Vocabulary, Comprehension, Motivation, and Special needs.

Each chapter had theoretical as well as practical (didactical) content. Relevant concepts, theories, and empirical findings were presented and discussed and at the end of each chapter at least two examples of classroom instruction were provided. Teachers were encouraged to use their own experience and adapt the reading instruction to their own classroom. Thus, they were not provided with a detailed manual but rather a set of educational tools to choose from and be inspired by.

Results
The results on all tests on each test occasion (T1-T3) are presented in Table 1. Due to the limited number of schools, especially in the FA only group, we decided to use ANOVAs to analyze main effects and interaction effects. Within-group Cohen’s $d$ was used to analyze effect sizes since there were some differences between groups at T1. For phonological awareness and letter-sound knowledge, ceiling effects were obtained already at the first test session and therefore results are not presented for T2 and T3 for these two variables.
Table 1.
Reading skills at T1-T3 for the three groups of children in grade 1 (M and SD).

<table>
<thead>
<tr>
<th></th>
<th>Full LegiLexi (n=217)</th>
<th>LegiLexi FA only (n=86)</th>
<th>Control (n=208)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>d</td>
</tr>
<tr>
<td>Pre-reading skills</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological awareness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 September</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letter-sound knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 September</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word decoding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word reading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 September</td>
<td>14.90</td>
<td>17.31</td>
<td>17.76</td>
</tr>
<tr>
<td>T2 December</td>
<td>33.02</td>
<td>20.16</td>
<td>32.95</td>
</tr>
<tr>
<td>T3 May</td>
<td>51.11</td>
<td>22.85</td>
<td>49.11</td>
</tr>
<tr>
<td>Cohen's d¹</td>
<td>1.79</td>
<td></td>
<td>1.33</td>
</tr>
<tr>
<td>Pseudo word reading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 September</td>
<td>5.97</td>
<td>6.47</td>
<td>7.73</td>
</tr>
<tr>
<td>T2 December</td>
<td>12.08</td>
<td>7.11</td>
<td>12.73</td>
</tr>
<tr>
<td>T3 May</td>
<td>17.49</td>
<td>8.08</td>
<td>18.11</td>
</tr>
<tr>
<td>Cohen's d¹</td>
<td>1.57</td>
<td></td>
<td>1.10</td>
</tr>
<tr>
<td>Language comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 September</td>
<td>15.49</td>
<td>3.58</td>
<td>15.23</td>
</tr>
<tr>
<td>T2 December</td>
<td>16.61</td>
<td>2.83</td>
<td>17.30</td>
</tr>
<tr>
<td>T3 May</td>
<td>17.27</td>
<td>3.17</td>
<td>18.09</td>
</tr>
<tr>
<td>Cohen's d¹</td>
<td>0.53</td>
<td></td>
<td>0.95</td>
</tr>
<tr>
<td>Listening comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sentences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 September</td>
<td>9.27</td>
<td>2.24</td>
<td>8.95</td>
</tr>
<tr>
<td>T2 December</td>
<td>10.24</td>
<td>1.87</td>
<td>9.81</td>
</tr>
<tr>
<td>T3 May</td>
<td>10.57</td>
<td>1.90</td>
<td>10.56</td>
</tr>
<tr>
<td>Cohen's d¹</td>
<td>0.63</td>
<td></td>
<td>0.85</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pictures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 September</td>
<td>2.20</td>
<td>2.83</td>
<td>3.28</td>
</tr>
<tr>
<td>T2 December</td>
<td>5.91</td>
<td>3.38</td>
<td>5.69</td>
</tr>
<tr>
<td>T3 May</td>
<td>8.97</td>
<td>3.13</td>
<td>8.13</td>
</tr>
<tr>
<td>Cohen's d¹</td>
<td>2.27</td>
<td></td>
<td>1.31</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>text</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 September</td>
<td>1.66</td>
<td>2.59</td>
<td>1.79</td>
</tr>
</tbody>
</table>
Two ANOVAs showed that there were no significant main effects of group at T1 for phonological awareness or letter-sound knowledge. For word reading, a repeated measures ANOVA revealed a significant interaction between test session and group ($p<.01$). The Full LegiLexi group improved their word reading the most between T1 and T3 (Cohen’s $d=1.79$ compared to 1.33 and 1.27). No significant interaction was found for pseudo word reading. Due to attrition following the randomization process, initial means on T1 differed between groups and the Full LegiLexi group had the lowest initial scores.

A repeated measures ANOVA showed that there was a significant interaction between test session and group for listening comprehension-words ($p<.01$). The group Legilexi FA only showed the biggest improvement between T1 and T3 (Cohen’s $d=0.95$ compared to 0.53 and 0.53). Effect sizes were smaller than for word reading. There was no significant interaction for listening comprehension-sentences.

There was a significant interaction between test session and group both for reading comprehension-pictures ($p<.01$) and reading comprehension-text ($p<.05$). For both reading comprehension tests, the Full LegiLexi group improved the most between T1 and T3. Cohen’s $d$ was 2.27 for Reading comprehension pictures, representing the biggest obtained effect size, compared to 1.31 and 1.35. For Reading comprehension text, Cohen’s $d$ was 1.75 compared to 1.45 and 1.16. Due to attrition following the randomization process, initial means on T1 differed between groups and the Full LegiLexi group had the lowest initial scores.

Discussion

The results demonstrate positive effects of participating in LegiLexi, especially for the group who had full access to both the formative assessment tool and the teacher course.
This group improved their word reading and reading comprehension the most between the first and third test session and effect sizes were substantial. There are two main explanations for these results.

First, it is possible that the improvements reflect actual positive effects on reading skills from using the formative assessment tool combined with the teacher course. This interpretation of results is in line with studies showing positive effects of teachers having access to student data (Connor et al., 2009). Some support for this explanation, regarding improvement in word reading for the full LegiLexi group, is that most teachers received recommendations to focus on decoding rather than language comprehension for the participating children. It is possible that these recommendations influenced their teaching in a way that was beneficial for student’s learning since students were in an early stage of their reading development. Furthermore, teachers in the full LegiLexi group had access to course material which included theoretical as well as practical, educational implementations to assist them in teaching word decoding. It is likely that an improvement in word reading skills also had positive effects on reading comprehension, since it is highly dependent on word decoding in early reading development (Catts, Hogan, & Adolf, 2005).

It should be acknowledged that there is an alternative explanation of the positive results obtained for the full LegiLexi group, namely regression to the mean. Even though groups were initially matched when they were assigned to conditions, due to attrition during the project period, the full LegiLexi group had somewhat lower initial scores on the reading tests. In general, participants with lower initial scores tend to score higher and closer to the mean when tests are repeated. However, it should be noted that when it comes to reading skills it is also quite possible that children with low initial scores are more difficult to remediate and often the gap between advanced and poor readers tend to grow over time (Stanovich, 1986).

A limitation of the present study is that test-retest correlations were only moderate for some of the tests used. The intervals between tests were several months and that
could to some extent explain the obtained correlations, but reliability and validity of tests should be focused in future work within the LegiLexi-project.

Thus, the results suggest that there were positive effects on reading skills for the LegiLexi-groups, especially the full LegiLexi group, but that results should be interpreted with some caution.

The results on language comprehension – words, where LegiLexi FA only showed the highest improvements, were somewhat surprising. Perhaps, teachers who did not have access to the course material spent more time on educational activities involving language comprehension and vocabulary than the full LegiLexi group, since this is common practice in Swedish schools, and increased time on task led to higher improvements. The difference compared to the control condition might reflect positive effects of using the formative assessment tool and adapting instruction according to recommendations for children with limited language comprehension skills relative to their other skills.

In line with RTI, LegiLexi provides a systematic use of tests which are repeated and where results are continuously documented. In LegiLexi this is followed by suggested adaptations of instruction, providing support for teachers’ decision making but still allowing for teachers to use their competence when planning and implementing reading instruction. Thus, LegiLexi is more flexible than a strict standard protocol version of RTI and more comparable to a problem solving approach (Berkley, Bender, Peaster, & Saunders, 2009). Compared to Formative assessment, pupils’ involvement in their learning processes and feedback to pupils are not explicitly focused. This can be regarded as a limitation of the current version of LegiLexi. On the other hand, feedback to teachers is central in LegiLexi and in line with formative assessment, assessments are forward-looking and related to learning goals (Black & William, 1998, 2018).

The results presented here stem from the pilot-year of LegiLexi and some changes will be made in the project regarding both content and form. The teacher course and information library will continue to be developed and refined. The same holds for the test battery and the recommendations for instruction. For example, a few items in a
couple of the tests will be replaced and additional versions of tests will be developed to reduce floor effects in grade 1 and ceiling effects in grade 3. This development of the project will involve close cooperation between teachers and researchers regarding the implementation of the different parts of LegiLexi in classroom practice. Qualitative data based on interviews with participating teachers will also be analyzed and will complement the quantitative data presented in this article. Thus, continuous adaptations will be made also on the project level, based on data and the experiences of participants.

The use of the LegiLexi assessment tool as such may contribute to the evidence base for the method as it is possible to collect all data that is generated in the testing procedures (pending consent from schools and parents). Hence, this study is possible to repeat on an even larger cohort once the general use has commenced.

Declaration of interest:

LegiLexi is a pro bono foundation of which XX is a cofounder. The assessment tool is provided free of charge to the school system. XX has not and will not receive any financial benefit from LegiLexi.

References

Anonymized authors, 2011
Anonymized authors, 2014
Anonymized authors, 2014
Anonymized authors, 2015


