The handle http://hdl.handle.net/1887/67090 holds various files of this Leiden University dissertation.

Author: Houwelingen, M.J. van
Title: Measurement numeracy education for prospective elementary school teachers: effects of inductive and deductive teaching on classroom interaction and student performance
Issue Date: 2018-11-27
References


Expertgroep Doorlopende Leerlijnen (2008). *Over de drempels met taal en rekenen [crossing boundaries with language and mathematics]*. Enschede: SLO.


Rasbash, J., Steele, F., Browne, W., & Prosser, B. (2015). A user’s guide to MLwiN. *Centre for Multilevel Modelling, University of Bristol, UK*.


List of Figures
2.1. Number of students in the whole sample, split by pretest participation, condition, and posttest participation (lesson attendance in the last column) ................................................................. 16
2.2. Category centroids on CATPCA dimensions 1 and 2 (N=257) .......................................................... 20
2.3. Students’ standardized scores on the CATPCA dimensions 2 and 3 (N=257) ................................. 21
2.4. Category centroids of active variables on the CATPCA dimensions 2 and 3 (N=257) ............... 23
2.5. Category centroids of supplementary variables on the CATPCA dimensions 2 and 3 (N=257) ... 25
3.1. Structural model of measurement numeracy in the main study, with factor correlations ........ 30
5.1. Quarter means (averaged over lessons) for question type per didactic approach ................... 67
5.2. Quarter means (averaged over lessons) for classroom interaction time per didactic approach .... 68
5.3. Means for question type per teacher (averaged over lessons and conditions) ......................... 69
5.4. Means for classroom interaction time per teacher (averaged over lessons and conditions) ..... 70
5.5. Lesson means for question type ........................................................................................................ 73
5.6. Quarter means for question type ....................................................................................................... 73
5.7. Lesson means for classroom interaction time .................................................................................... 75
5.8. Quarter means for classroom interaction time (averaged over lessons) .................................. 75
5.9. Lesson means for classroom interaction time per didactic approach ......................................... 76
6.1. Distribution of the pretest scores of the three aspects and the WISCAT score ......................... 88
6.2. Mean scores on posttest by teacher and condition for metric (top), scale (middle), and area (bottom) .......................................................................................................................... 91
List of Tables

2.1. Number of students (percentage of males) in pretest and posttest in 2013-2014 and 2014-2015; Mean age in years. ................................................................. 13
2.2. Number of students who participated (or not) in the pretest and the posttest. ................................................................. 14
2.3. Number of students (number of males in parentheses) in each student group, for the 153 students participating in both the pretest and the posttest (left), and for the whole sample of 259 students (right). ........................................................................... 15
2.4. Percentage of students attending class, per lesson, for the 153 students in both the pretest and the posttest (left), and for the whole sample of 259 students (right). ................................................................. 15
2.5. Percentage of males, mean age, percentage immigrants, and highest previous education for the 153 students who participated in both the pretest and the posttest (left), and the whole sample of 259 students (middle), compared to national numbers in 2014-2015 (right) (Central Bureau for Statistics, 2018). ......................................................................................... 17
2.6. Mean WISCAT score of students in the Netherlands participating in the WISCAT test for the first time, for the 153 students who participated in both the pretest and the posttest (left), and the whole sample of 259 students (middle), compared to national mean between 2007 and 2012 (right) (Eggen & Straetmans, 2013). ......................................................................................... 17
2.7. Percentage of students’ and their parents’ highest previous education, for the 153 students who participated in both the pretest and the posttest (left), and the whole sample of 259 students (right). 18
2.8. Number of students in the sample who scored sufficiently at mathematics in their previous education, compared to students who did not follow mathematics classes or scored insufficiently, for the 153 students who participated in both the pretest and the posttest (left), and the whole sample of 259 students (right). ......................................................................................... 18
2.9. Mean pretest score and mean WISCAT score of different student categories. ................................................................. 24
3.1. Item examples per hypothesized sub-skill................................................................................................................................. 31
3.2. Model fit of 3-factor models and 1-factor models for the pilot study pretest (without estimating measures items), and difftest to determine if a 3-factor model fits significantly better than a 1-factor model ................................................................................................................................. 34
3.3. Model fit of 3-factor models and 1-factor models for the pilot study posttest (without estimating measures items), and difftest to determine if a 3-factor model fits significantly better than a 1-factor model ................................................................................................................................. 34
3.4. Model fit of 3-factor models and 1-factor models for the pretest and posttest of the main study, and difftest to determine whether or not a 3-factor model fits significantly better than a 1-factor model ................................................................................................................................. 35
3.5. Factor correlations of the main study 3-factor model ................................................................................................................................. 35
3.6. Factor loadings main study pretest 3-factor model ................................................................................................................................. 35
3.7. Factor loadings main study posttest 3-factor model ................................................................................................................................. 37
3.8. Reliability of the main study item sets (10 items for metric, scale and area sub-skill). ................................................................. 38
4.1. Example exercises for each aspect ......................................................................................................................................................... 46
4.2. Teacher’s didactic approach preference per mathematical aspect ................................................................................................................................. 49
4.3. Focus group advice for the deductive approach ......................................................................................................................................................... 50
4.4. Focus group advice for the inductive approach ......................................................................................................................................................... 52
5.1. Distribution of student groups to teachers and teaching approach, with initial number of students per group, percentage male students, and number of students present per lesson ................................................................................................................................. 52
5.2. Intraclass correlation coefficient – interrater reliability for classroom interaction time ......................................................................................................................................................... 62
5.3. Consistency between coders ......................................................................................................................................................... 63

121
5.4. Consistency between coders for question type coding – Cohen’s Kappa per approach and for totals. .................................................................................................................................................. 64
5.5. Percentage of teacher behavior as instructed (fidelity check). ................................................................................................................................. 64
5.6. Distribution of slots to quarters, for different numbers of coded slots per lesson. ............................................................................................................ 65
5.7. Mean question type (0=controlling, 1=same, 2=stimulating) for conditions. ............................................................................................................ 66
5.8. Mean classroom interaction time (between 0 and 120 seconds) for conditions. ............................................................................................................ 66
5.9. Between-groups effects on the teachers’ question type............................................................................................................................ 66
5.10. Between-groups effects on the classroom interaction time............................................................................................................................ 66
5.11. Mean question type (0=controlling, 1=same, 2=stimulating) for teachers by condition........................................................................... 70
5.12. Mean classroom interaction time (between 0 and 120 seconds) for teachers by condition........................................................................... 71
5.13. Within-groups effects on teacher’s question type................................................................................................................................. 72
5.14. Within-groups effects on the classroom interaction time............................................................................................................................ 74
6.1. Item examples per aspect................................................................................................................................. 81
6.2. Multilevel check: -2*loglikelihood for the four models, for three different aspects. .................................................. 83
6.3. Independent t-test for pretest score differences in ten pre-existing groups (N_{deductive}=80, N_{inductive}=73). .................................................. 84
6.4. Independent t-test for WISCAT score differences in ten pre-existing groups (N_{deductive}=76, N_{inductive}=67). .................................................. 84
6.5. Mean WISCAT scores per teacher. ................................................................................................................................. 85
6.6. One-way ANOVA to check for significant differences in WISCAT score per teacher........................................................................... 85
6.7. Differences per teacher in pretest score for the aspects metric, scale, and area........................................................................... 85
6.8. One-way ANOVA to check for significant differences per teacher in pretest score for the aspects metric, scale, and area........................................................................... 85
6.9. Spearman’s rho correlation coefficients between pretest scores and WISCAT score (1-tailed)... 87
6.10. Normality (kurtosis and skewness) of the three pretest scores and the WISCAT score. ................................. 87
6.11. Mean pretest score and mean posttest score per aspect (N=153). ..................................................................................... 89
6.12. Mean posttest scores for metric, scale, and area, by teacher by condition........................................................................... 90
6.13. Results of student questionnaires in the deductive approach (drawn after each lesson). ........................................... 92
6.14. Results of student questionnaires in the inductive approach (drawn after each lesson). ........................................... 92
6.15. Effect on posttest score for metric, scale, and area........................................................................... 93
6.16. Model3 effect sizes (partial eta squared) and regression weights, of pretest and WISCAT, for metric, scale, and area........................................................................... 94
### Appendix

Table A.1. Factor loadings pilot study pretest for subtests A1, A2, B1, B2, C1, and C2.

<table>
<thead>
<tr>
<th>factor</th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.07</td>
<td>.12</td>
<td>.35</td>
<td>.37</td>
<td>.15</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.31</td>
<td>.80</td>
<td>.38</td>
<td>.67</td>
<td>.13</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.20</td>
<td>.41</td>
<td>.36</td>
<td>.28</td>
<td>-.06</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>-.16</td>
<td>.26</td>
<td>.44</td>
<td>.04</td>
<td>.42</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1.23</td>
<td>.67</td>
<td>.35</td>
<td>.64</td>
<td>.65</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>.69</td>
<td>.38</td>
<td>.11</td>
<td>.11</td>
<td>.54</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>.59</td>
<td>.70</td>
<td>.37</td>
<td>.83</td>
<td>.81</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>.58</td>
<td>.61</td>
<td>.68</td>
<td>.57</td>
<td>.28</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>.35</td>
<td>.55</td>
<td>-.02</td>
<td>.44</td>
<td>.51</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>.90</td>
<td>.91</td>
<td>.47</td>
<td>.78</td>
<td>.79</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>.59</td>
<td>.95</td>
<td>.95</td>
<td>.81</td>
<td>.86</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>.80</td>
<td>.90</td>
<td>.93</td>
<td>.67</td>
<td>.96</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>.62</td>
<td>.94</td>
<td>.51</td>
<td>.75</td>
<td>.92</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>.94</td>
<td>.87</td>
<td>.83</td>
<td>.74</td>
<td>.91</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>.84</td>
<td>.99</td>
<td>.80</td>
<td>.86</td>
<td>.82</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>.60</td>
<td>.89</td>
<td>.54</td>
<td>.91</td>
<td>.89</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>.31</td>
<td></td>
<td></td>
<td></td>
<td>.49</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>.45</td>
<td>.89</td>
<td>.82</td>
<td>.72</td>
<td>.91</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>.50</td>
<td>.41</td>
<td>.77</td>
<td>.68</td>
<td>.46</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>.84</td>
<td>.66</td>
<td>.61</td>
<td>.83</td>
<td>.93</td>
</tr>
<tr>
<td>21</td>
<td>4</td>
<td>.77</td>
<td>.66</td>
<td>.70</td>
<td>.91</td>
<td>.83</td>
</tr>
<tr>
<td>22</td>
<td>4</td>
<td>.61</td>
<td>.51</td>
<td>.92</td>
<td>.83</td>
<td>.68</td>
</tr>
<tr>
<td>23</td>
<td>4</td>
<td>.44</td>
<td>.80</td>
<td>.80</td>
<td>.44</td>
<td>.77</td>
</tr>
<tr>
<td>24</td>
<td>4</td>
<td>.64</td>
<td>.40</td>
<td>.74</td>
<td>.78</td>
<td>.66</td>
</tr>
</tbody>
</table>

Note: This table contains factor loadings of 144 unique items (6 subtests times 24 items: items A1_1 through A1_24, A2_1 through A2_24, et cetera).

Note: items were designed to measure a specific sub-skill: estimating measures (items 1-6, factor 1), understanding relationships within the metric system (items 7-12, factor 2), calculating with scale (items 13-18, factor 3), and calculating length, area and volume (items 19-24, factor 4).

Note: the model for subtest B2 could not be identified.

Note: a blank spot indicates that none of the students answered that item correctly.

Note: non-significant factor loadings are marked in italics.
Table A.2. Factor loadings pilot study posttest for subtests A1,A2,B1,B2,C1, and C2.

<table>
<thead>
<tr>
<th>factor</th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.11</td>
<td>.34</td>
<td>.57</td>
<td>.39</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.24</td>
<td>.47</td>
<td>.23</td>
<td>.06</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.50</td>
<td>.51</td>
<td>.18</td>
<td>.48</td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-.12</td>
<td>.19</td>
<td>.24</td>
<td>.27</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.80</td>
<td>.13</td>
<td>.49</td>
<td>.77</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.26</td>
<td>.74</td>
<td>.00</td>
<td>.38</td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.06</td>
<td>.48</td>
<td>.61</td>
<td>.91</td>
<td>.66</td>
<td>.58</td>
</tr>
<tr>
<td>8</td>
<td>.56</td>
<td>.87</td>
<td>.34</td>
<td>.62</td>
<td>.76</td>
<td>.23</td>
</tr>
<tr>
<td>9</td>
<td>.39</td>
<td>.57</td>
<td>.66</td>
<td>.76</td>
<td>.52</td>
<td>.48</td>
</tr>
<tr>
<td>10</td>
<td>.47</td>
<td>.69</td>
<td>.77</td>
<td>.77</td>
<td>.64</td>
<td>.76</td>
</tr>
<tr>
<td>11</td>
<td>.51</td>
<td>.55</td>
<td>.93</td>
<td>.80</td>
<td>.72</td>
<td>.61</td>
</tr>
<tr>
<td>12</td>
<td>.75</td>
<td>.50</td>
<td>.94</td>
<td>1.01</td>
<td>.43</td>
<td>.81</td>
</tr>
<tr>
<td>13</td>
<td>.61</td>
<td>.84</td>
<td>.98</td>
<td>.49</td>
<td>.73</td>
<td>.56</td>
</tr>
<tr>
<td>14</td>
<td>.73</td>
<td>.49</td>
<td>.48</td>
<td>.44</td>
<td>.60</td>
<td>.49</td>
</tr>
<tr>
<td>15</td>
<td>.55</td>
<td>.59</td>
<td>.79</td>
<td>.59</td>
<td>.47</td>
<td>.79</td>
</tr>
<tr>
<td>16</td>
<td>.63</td>
<td>.87</td>
<td>.45</td>
<td>.72</td>
<td>.76</td>
<td>.76</td>
</tr>
<tr>
<td>17</td>
<td>-.35</td>
<td>-.35</td>
<td></td>
<td></td>
<td></td>
<td>.72</td>
</tr>
<tr>
<td>18</td>
<td>.34</td>
<td>.76</td>
<td>.48</td>
<td>.25</td>
<td>.83</td>
<td>.70</td>
</tr>
<tr>
<td>19</td>
<td>.24</td>
<td>.34</td>
<td>.28</td>
<td>.48</td>
<td>.42</td>
<td>.10</td>
</tr>
<tr>
<td>20</td>
<td>.37</td>
<td>.46</td>
<td>.78</td>
<td>.78</td>
<td>.40</td>
<td>.61</td>
</tr>
<tr>
<td>21</td>
<td>.36</td>
<td>.76</td>
<td>.87</td>
<td>.82</td>
<td>.75</td>
<td>1.00</td>
</tr>
<tr>
<td>22</td>
<td>.57</td>
<td>.64</td>
<td>.58</td>
<td>.59</td>
<td>.89</td>
<td>.70</td>
</tr>
<tr>
<td>23</td>
<td>.77</td>
<td>.83</td>
<td>.52</td>
<td>.86</td>
<td>.33</td>
<td>.60</td>
</tr>
<tr>
<td>24</td>
<td>.91</td>
<td>.62</td>
<td>.27</td>
<td>.72</td>
<td>.61</td>
<td>.60</td>
</tr>
</tbody>
</table>

Note: This table contains factor loadings of 144 unique items (6 subtests times 24 items: items A1_1 through A1_24, A2_1 through A2_24, et cetera).

Note: items were designed to measure a specific sub-skill: estimating measures (items 1-6, factor 1), understanding relationships within the metric system (items 7-12, factor 2), calculating with scale (items 13-18, factor 3), and calculating length, area and volume (items 19-24, factor 4).

Note: the model for subtest B2 could not be identified.

Note: a blank spot indicates that none of the students answered that item correctly.

Note: non-significant factor loadings are marked in italics.
Table A.3. *Independent samples T-test: effect of student characteristics on pretest score.*

<table>
<thead>
<tr>
<th>metric</th>
<th>home language</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>df</td>
<td>p</td>
<td>r</td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td>1.907</td>
<td>208</td>
<td>0.058</td>
<td>0.131</td>
<td></td>
</tr>
<tr>
<td>mathematical history</td>
<td>-3.931</td>
<td>145</td>
<td>0.000</td>
<td>0.310</td>
<td></td>
</tr>
<tr>
<td>previous education</td>
<td>-2.852</td>
<td>208</td>
<td>0.005</td>
<td>0.194</td>
<td></td>
</tr>
<tr>
<td>education mother</td>
<td>0.389</td>
<td>127</td>
<td>0.698</td>
<td>0.194</td>
<td></td>
</tr>
<tr>
<td>education father</td>
<td>-0.518</td>
<td>122</td>
<td>0.605</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>metric</th>
<th>home language</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>df</td>
<td>p</td>
<td>r</td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td>4.269</td>
<td>83.480</td>
<td>0.000</td>
<td>0.423</td>
<td></td>
</tr>
<tr>
<td>mathematical history</td>
<td>-4.133</td>
<td>145</td>
<td>0.000</td>
<td>0.325</td>
<td></td>
</tr>
<tr>
<td>previous education</td>
<td>-3.113</td>
<td>208</td>
<td>0.002</td>
<td>0.211</td>
<td></td>
</tr>
<tr>
<td>education mother</td>
<td>-0.088</td>
<td>127</td>
<td>0.930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>education father</td>
<td>-1.999</td>
<td>122</td>
<td>0.048</td>
<td>0.178</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>metric</th>
<th>home language</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>df</td>
<td>p</td>
<td>r</td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td>3.169</td>
<td>208</td>
<td>0.002</td>
<td>0.215</td>
<td></td>
</tr>
<tr>
<td>mathematical history</td>
<td>-4.047</td>
<td>144.967</td>
<td>0.000</td>
<td>0.319</td>
<td></td>
</tr>
<tr>
<td>previous education</td>
<td>-3.848</td>
<td>202.630</td>
<td>0.000</td>
<td>0.261</td>
<td></td>
</tr>
<tr>
<td>education mother</td>
<td>-0.327</td>
<td>127</td>
<td>0.724</td>
<td></td>
<td></td>
</tr>
<tr>
<td>education father</td>
<td>-1.928</td>
<td>122</td>
<td>0.056</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>