

University of Groningen

A captivating snapshot of standardized testing in early childhood

Frans, Niek

DOI:
[10.33612/diss.95431744](https://doi.org/10.33612/diss.95431744)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Frans, N. (2019). *A captivating snapshot of standardized testing in early childhood: on the stability and utility of the Cito preschool/kindergarten tests*. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen. <https://doi.org/10.33612/diss.95431744>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Appendix A: Supplementary information Chapter 2

Table A1

Exploratory Mokken scale analysis with item H coefficients for the two scales

	Brown (2006) factor structure [†]		H_i
	Second order	First order	
<i>Scale: The Cito test is useful for teachers (H = .34)</i>			
Assessment results should be treated cautiously because of measurement error*	Irrelevant	Inaccurate	.38
Assessment forces teachers to teach in a way against their beliefs*	Irrelevant	Bad	.37
Teachers should take into account the error and imprecision in all assessment*	Irrelevant	Inaccurate	.36
Assessment results are filed and ignored*	Irrelevant	Ignored	.30
Teachers conduct assessment but make little use of the results*	Irrelevant	Ignored	.29
<i>Scale: The Cito test provides valid information (H = .40)</i>			
Assessment provides information on how well teachers are doing	.	.	.50
Assessment helps students improve their learning	Improvement	Improves learning	.48
Assessment is a way to determine how much students have learned...	Improvement	Describes abilities	.48
Assessment is a good way to evaluate a school	School Accounting	-	.48
Assessment is an accurate indicator of a school's quality	School Accounting	-	.46
Assessment establishes what students have learned	Improvement	Describes abilities	.45
Assessment results can be depended on	Improvement	Valid	.45
Assessment is an accurate indicator of a teacher's quality	.	.	.45
Assessment results are trustworthy	Improvement	Valid	.43
Assessment feeds back to students their learning needs	Improvement	Improves learning	.42
Assessment results are consistent	Improvement	Valid	.42
Assessment is integrated with teaching practice	Improvement	Improves teaching	.42
Assessment determines if students meet qualification standards	Student Accounting	-	.42
Assessment is a good way to evaluate a teacher	.	.	.42
Assessment provides information on how well a group is doing	.	.	.42
Assessment is an imprecise process*	Irrelevant	Inaccurate	.41
Assessment provides information on how well schools are doing	School Accounting	-	.38
Assessment is unfair to students*	Irrelevant	Bad	.35
Assessment measures students' higher order thinking skills	Improvement	Describes abilities	.34
Assessment information modifies ongoing teaching of students	Improvement	Improves teaching	.33
Assessment allows different students to get different instruction	Improvement	Improves teaching	.33
Assessment provides feedback to students about their performance	Improvement	Improves learning	.30
Assessment interferes with teaching*	Irrelevant	Bad	.30
Assessment is assigning a grade or level to student work	Student Accounting	-	.25

Note: The items 'Assessment places students into categories' and 'Assessment has little impact on teaching' did not fit any scale. Reverse-coded items are indicated with *. [†]These two columns indicate the original factors in the CoA-III as found by Brown (2006). Items added to the questionnaire are indicated by a '.' in the original factor. Removing these items had no effect on the scales.

Appendix

Table A2
Full codebook of the interview data

Coding theme	Main codes (number of subcodes, if applicable)	Example
Necessary conditions for testing	Conditions for testing related to the child (7) Practical conditions for testing Conditions related to the teacher (2)	The child needs to be able to focus The test takes a lot of time You need to know the manual somewhat
Strategies to accommodate conditions	Before test administration During test administration After test administration	We avoid the word 'test' Children that have trouble concentrating sit close to me If a child is anxious I re-test him or her one-on-one
Target group for test administration	Dependent on the grade-level Dependent on the previous test score Dependent on grade retention Dependent on confidence teacher Dependent on parent request	We don't administer the test in preschool We re-test children in June if they score a D/E in January We don't test a child that is going to repeat kindergarten I only test a child when I have doubts about his/her level I sometimes re-test when the parents ask me to
Emotionally charged statement	Positive affect teacher Negative affect teacher Positive affect child Negative affect child Positive affect other stakeholder Negative affect other stakeholder	I'm glad that we have this test It's a horrible test Children love working in a booklet Children get stressed and anxious when tested Parents think the test is important Some of my colleagues hate these tests
Relationship to the curriculum	Play should be central in the curriculum Cognitive challenge is important in (pre-)K Education should be child directed The test is not on (pre-)K level Other skills than those tested are important	Children learn mainly by playing Challenging children in language is vital at a young age Children will ask about writing when they are ready The level of the test is too difficult for many children His score is good but he still acts too young for his age
Information gained from the test	Achievement level Differences in test scores over time Scores on subcategories Observation during the test administration Answers to specific questions	This child scored a D on language You can see that her achievement score has gone up If number sense is low, you can focus on that I noticed that he is unable to listen to my instructions You can see here that he worked from right to left
Alternative means to the test	(Un)structured observation by the teacher Teacher designed tests Other external tests	You have your own observations which tell you a lot I gave them a small task to see if they could do it We also have a vocabulary test in October
Professional autonomy of teachers	Teacher's sense of trust Teacher's sense of professionalism Teacher's sense of freedom and pressure Teacher's sense of autonomy support	Why do you need the test, trust the teacher for once And it's like someone wants to check if I'm good enough I am forced to administer this test Sometimes you miss things that the test helps you see
Purpose according to the teacher	Confirmation for the teacher/child (2) Evaluation of the teacher/curriculum (1) Evaluation of the child's mastery (3) Guideline for what a child is expected to know Guideline for what a teacher is expected to teach Familiarizing children with formal testing Indication for grade-skipping or retention	The test is a confirmation for you as a teacher The test can tell me if what I offered was sufficient The test shows what a child can and cannot do The test shows what a child needs to know I look at the test to see what needs to be taught For us the main idea is that children get used to testing I would be hesitant to send him to first grade with two D's
Expectations of other stakeholders	Control/confirmation of educational process Making the grade (scoring at least average) Growth between test administrations Few or no expectations	Parents want to know a child's level, the test provides this The school wants to know if I am on par with expectations Parents want to see if their child has grown The educational inspection is not interested in the results

Table A2
Continued

Use or impact of the test	Use of the results (7)	We use the results to arrange children into groups
	Impact of the test on education (2)	I teach the word 'Antlers' as they often struggle with it
	Impact of the test on the behavior of others	Parents practice at home so their child scores higher
	Limited impact or use of the test	In practice you don't do a lot with the results
Characteristics of the test	Form in which children are tested	Assignments in the test are all in 2D
	Content of the test	Questions can be interpreted in multiple ways
	Administration of the test (2)	The test is just a snapshot of the child's development
Societal context (of the child)	Continuity between tests	The tests between years are so different
	Background/context of the child	Some children just learn more from their parents
	Higher external demands	Society just expects more nowadays

Appendix B: Supplementary analyses Chapter 4 and 5

Table B1 shows the distribution on several demographics for the entire sample ($N = 1407$) of which subsamples were used in Chapters 4 and 5. When available, the proportion in the Dutch population as reported by the Central Bureau of Statistics and CPB is shown alongside the sample proportions. The table shows that the sample closely resembles the population on most variables. A notable exception is that our sample contains relatively few children with a foreign heritage.

Table B1

Sample descriptives student level, compared to population statistics.

	Sample proportion	Population proportion*
Sex [Female]	.50	.50
Linear school career [1-5]**	.88	.85
One parent household	.11	.14
Foreign Heritage	.09	.16
Special needs funding	.02	
Special school funding 0.3	.06	.07
Special school funding 1.2	.04	.05
Lateral entry	.13	
Had an IEP [grade 1-5]	.28	
Mean age July 2014 [years] (<i>SD</i>)	9.28 (0.47)	

*Proportions of Dutch youth in primary school 2011/2012 according to CBS Statline (March 2016).

The estimated proportion of children who had a linear school career is taken from CPB (2011/2012) over all children in primary education

**Pupils that are under the age of seven in October of first grade are never considered as 'repeating kindergarten'

Table B2 compares our sample on the variables reported by Cito to their norm sample and population values. The reported numbers are taken from the latest update of the norms for the Mathematics, Spelling and Reading comprehension tests in 2011/2012. Compared to the population, the sample is almost equally distributed on urbanization degree and sex, i.e. there are only small and non-significant differences between the sample and population distribution. On the other hand, there is a clear overrepresentation of schools in the Northern provinces, whereas the Western and Southern provinces are underrepresented. In addition, it appears that schools that contain relatively few [0%,10%) children from low-educated households are somewhat overrepresented. Then again, there is also an overrepresentation of schools where between 25% and 40% of the children in the sample come from low educated households.

Table B2

Sample proportions compared to distribution in Dutch primary school population and Cito calibration sample

	Variable	Sample students	Cito '11/ '12	Population '11/ '12	$\chi^2(df)$ Cito	φ Cito	$\chi^2(df)$ Pop.	φ Pop.
Urbanization	Rural	55.5	62.7	56.3	31.1 (1)*	.15	.3 (1)	.01
	Urban	44.5	37.3	43.7				
Sex	Male	49.7	50.6	50.4	0.5 (1)	.02	.3 (1)	.01
	Female	50.3	49.4	49.6				
Region	North	22.7	13.6	10.2	182.6 (3)*	.35	252.3 (3)*	.42
	East	23.4	24.7	22.7				
	West	38.9	33.5	47.1				
	South	15.1	28.1	20.0				
Low educ.	[0,.10)	70.2	63.5	60.6	125.6 (3)*	.30	128.9 (3)*	.31
	[.10, .25)	15.3	25.9	26.4				
	[.25, .40)	10.5	5.7	6.6				
	[.40, 1]	4.0	4.9	6.4				

Note: Significance at $\alpha = .05$ is indicated by an asterisk. The reported X^2 compare the sample distribution to the population distribution. The letter φ denotes the phi correlation coefficient.

Table B3 compares the sample distribution of children from low-educated households in more detail, with a dataset from DUO that contains 6901 primary schools. The table shows an underrepresentation of schools with the highest proportion of children from low-educated households (i.e. $\geq 40\%$) and an overrepresentation of schools with the lowest proportion of children from low educated households (i.e. $< 10\%$). Because the number of students within each school differ, the distribution of children in the sample is slightly different and is shown in the second half of the table.

Table B3

Percentage of children with low educated parents compared to population data (DUO, Oct. 2012)

Low educ.	School level					Child level				
	N DUO	% DUO	N sample	% sample	% Differ.	N DUO	% DUO	N sample	% sample	% Differ.
[0,.10)	4071	59.0	36	63.2	4.2	942985	63.0	955	70.2	7.2
[.10,.20)	1546	22.4	8	14.0	-8.4	309170	20.6	156	11.5	-9.1
[.20,.30)	537	7.8	7	14.0	6.2	99749	6.7	112	8.2	1.5
[.30,.40)	284	4.1	4	5.3	1.2	55672	3.7	83	6.1	2.4
[.40,.50)	217	3.1	1	1.8	-1.3	43281	2.9	41	3.0	.1
[.50,.60)	150	2.2	1	1.8	-.4	29003	1.9	13	1.0	-.9
[.60,.70)	61	.9	0	.0	-.9	11773	.8	0	.0	-.8
[.70,.80)	26	.4	0	.0	-.4	5242	.3	0	.0	-.3
[.80,.90)	8	.1	0	.0	-.1	869	.1	0	.0	-.1
[.90,1]	1	.0	0	.0	.0	88	.0	0	.0	.0
Total N	6901		57			1,5*10 ⁶		1360		

Appendix C: Supplementary information Chapter 5



Figure C1: Set of ten score cards presented to teachers who participated in the interviews of Chapter 2.

In the fourth interview of Chapter 2, teachers were asked to rank ten score cards (shown in Figure C1) from least alarming to most alarming. The presented cards were selected in such a way that different interpretations (e.g. focus on growth, focus on score magnitude) would lead to different rankings of the cards. Teachers were given the following instruction:

Finally I have a small task to end this interview. In front of you are the M2 and E2 scores of 10 hypothetical children. Each card represents the scores of one single child expressed in the categories I to V. You can assume that each child has a score in the middle of these categories, so there are no high V and low V scores etc. It is a

simple example, but can you tell me which pairs of scores you would be worried, a little worried or not worried at all? Next, can you rank these scores from least to most alarming?

Each of the cards in Figure C1 has an identification number between 1 and 10. The rankings of these pairs by each of the teachers in the interview study is presented in Table C1. Although not all teacher rankings corresponded perfectly with a specific rationale, three distinct approaches could be seen. The ranking of Ria, Rianne and Renee coincided almost completely (Spearman correlations of .98, .96 and .95 respectively) with a ranking that prioritizes the last score (lowest to highest) and subsequently ranks on growth (least to most). Mona ranks all the score pairs perfectly on their average. While Ina and Irina seem to rank mainly on growth ($r_s = .68$ and $r_s = .81$ respectively).

One notable exception is that all teachers indicated that cards 9 and 1 were very alarming, regardless if they ranked all other cards on their growth first. Teachers frequently expressed the idea that these children did not grow, even though these score pairs represent average growth. The task is used in this dissertation as an illustration of different score interpretations (Chapter 5).

Table C1

Score cards and ranking from least (1) to most (10) alarming by the different teachers in Chapter 2.

Score ID	Score cards			Teacher ranking					
	Score M2	Score E2	Growth in percentiles	Ria	Rianne	Ina	Irina	Renee	Mona
1	4	4	0	8	8	8	8.5	9	8
2	3	3	0	6	5	6	4	3	5
3	1	2	-20	2	2	5	5	1	1
4	1	3	-40	7	9	10	8.5	7	3
5	4	3	20	4	4	2	2	5	6
6	2	2	0	1	1	7	3	2	2
7	2	3	-20	5	6	4	8.5	6	4
8	5	3	40	3	3	1	1	4	7
9	5	5	0	10	10	9	8.5	10	10
10	5	4	20	9	7	3	6	8	9

Note: Irina ranked several cards the same. These cards all received the same average rank of 8.5.

