Three Studies on Learning to Learn in Finland : Anti-Flynn Effects 2001-2017

Vainikainen, Mari-Pauliina

2022

Vainikainen , M-P & Hautamäki , J 2022 , ' Three Studies on Learning to Learn in Finland : Anti-Flynn Effects 2001-2017 ' , Scandinavian Journal of Educational Research , vol. 66 , no. 1 , pp. 43-58 . https://doi.org/10.1080/00313831.2020.1833240

http://hdl.handle.net/10138/340684
https://doi.org/10.1080/00313831.2020.1833240

Scandinavian Journal of Educational Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/csje20

Three Studies on Learning to Learn in Finland: Anti-Flynn Effects 2001–2017

Mari-Pauliina Vainikainen & Jarkko Hautamäki

SCANDINAVIAN JOURNAL OF EDUCATIONAL RESEARCH
2022, VOL. 66, NO. 1, 43–58
https://doi.org/10.1080/00313831.2020.1833240

Three Studies on Learning to Learn in Finland: Anti-Flynn Effects 2001–2017

Mari-Pauliina Vainikainen [a,b] and Jarkko Hautamäki[b]

[a]Faculty of Education and Culture, Tampere University, Tampere, Finland; [b]Centre for Educational Assessment, University of Helsinki, Helsinki, Finland

ABSTRACT
Finland is known for its high-performing educational system, but local assessments have shown that performance has declined during the past decade. We report the results of nationally representative learning to learn assessments in which 15-year-olds took an identical test in the same schools in 2001, 2012 and 2017. The results show that the level of both domain-general cognitive performance and learning-related beliefs dropped dramatically from 2001 to 2012, but the negative trend has stopped since then. For learning-related beliefs, the 2017 results were approaching the 2001 baseline level. The findings indicate that we may not be dealing with a true anti-Flynn effect, but the decline can possibly be explained by reduced motivation and effort in low-stakes assessment and schoolwork.

ARTICLE HISTORY
Received 24 February 2020
Accepted 19 August 2020

KEYWORDS
Learning to learn; anti-Flynn effect; cognitive competences; learning-related beliefs; decline of test scores

Since early 2000, Finland has been in the spotlight due to its high-performing equitable education system (e.g., OECD, 2001; 2016). About ten years later, the first signs of decline emerged in a municipal learning to learn assessment study (Kupiainen et al., 2011). It led to the revival of a national learning to learn (L2L) assessment programme launched in the mid-1990s to monitor the development of trends in this area of twenty-first century skills (Hautamäki et al., 2013). This study compares the results of Finnish 15-year-olds in 2012 and 2017 to the baseline study conducted in 2001, covering the core areas of domain-general cognitive competences and mastery and detrimental learning-related beliefs of the Finnish L2L framework (Hautamäki et al., 2002). As there was a large body of anchor items in all assessment cycles and the same schools participated in each cycle, we fit confirmatory factor models on three nationally representative random samples (Total $N = 9746$) of ninth graders. The aim is to confirm the locally observed decline of early 2010s with national data and to reveal the most recent developments.

Assessing Learning to Learn

Learning to learn (L2L) was included in the Finnish educational assessment plan (National Board of Education, 1999) in the mid-1990s as a result of a worldwide interest in the measurement of cross-curricular competences (see Voogt & Roblin, 2012, for an overview). L2L was defined as cognitive competences that are needed in all learning and the willingness to use these competences (Hautamäki et al., 2002). The measurement is based on domain-general cognitive tasks, questionnaires about learning-related beliefs, and more recently, log data analysis of task behaviour. The Finnish

CONTACT Mari-Pauliina Vainikainen mari-pauliina.vainikainen@tuni.fi Faculty of Education and Culture, Tampere University, Åkerlundinkatu 5, 33014, Tampere, Finland

L2L model is close to many definitions of twenty-first century skills that have a central role in educational policy discussions all over the world (Griffin & Care, 2015; Lai & Viering, 2012; Pellegrino & Hilton, 2012; Van Laar et al., 2017; Voogt & Roblin, 2012).

The origin of the concept L2L can be found already in Harlow's studies (1949), and it was also used in other contexts (e.g., Cronbach & Snow, 1977; Pask, 1976) before it was adopted in educational policy documents in the mid-1990s. The concept was introduced in European Union discussion papers, first under cross-curricular skills, then in the list of key competencies (for the history, see Hautamäki et al., 2010). As a concept, it allowed for semantic interpretations within the organisation system of education systems (Baraldi & Corsi, 2017). It was included, in differing ways, in national documents on educational assessment and evaluation (e.g., Fredriksson & Hoskins, 2007; Habók, 2015; Hoskins & Deakin Crick, 2010; Moreno & Martín, 2007; Visentin, 2017).

The Finnish work on L2L started in 1996 based on the general assessment framework of the National Board of Education (English version published in 1999), leading to the publication of the Finnish L2L framework in 2002 (Hautamäki et al., 2002). The definition is twofold: L2L is the ability and willingness to start to work, to commit oneself to intellectually demanding tasks, which are classified into reasoning, mathematical and reading domains of cognition (see Carroll, 1963; Sternberg et al., 2001) in order to represent the domain-general ability estimated to be needed in the twenty-first century (Adey et al., 2007; Greiff et al., 2014; Greiff et al., 2015; Vainikainen et al., 2015). The affective component – controlling the willingness to engage with new tasks with unknown a priori success rate, but which always have some well-known elements like numbers and words – is measured with several Likert self-report scales. In this paper, they are reported as two higher-order factors: learning-related *mastery* and *detrimental* beliefs. The naming of these scales is related to Bandura's self-efficacy theory (1997) and the self-concepts approach of Marsh et al. (1988), and reflects the classical two principal forces – vectors – of motivation (Atkinson, 1964: hope of success and fear of failure). The items used in measuring these two broader constructs have their roots in achievement goal theory (achievement and avoidance orientation; Harackiewicz et al., 2002) self-handicapping beliefs (Urdan & Midgley, 2001) and causal beliefs about the role of luck in explaining success as measured within the framework of action control theory (Little et al., 2001). Besides achievement orientation, the mastery beliefs factor comprised a scale for measuring students' attitudes towards school, which was developed for the original Finnish L2L test in the mid-1990s and which have been then validated in multiple large-scale studies over decades (e.g., Kupiainen et al., 2014).

Anti-Flynn Effect

The anti-Flynn effect refers to a decline in cognitive performance at the population level. After decades of slight but steady improvement in intelligence scores during the twentieth century (Flynn, 1984), the trend seems to have reversed in at least some European countries (Flynn & Shayer, 2018). Flynn and Shayer (2018) reviewed studies on anti-Flynn effects with varying assessments (Raven, WISC and WAIS, military factor test batteries, and Piagetian instruments), with varying age groups (14- to 16-year olds, transitional ages over 17 up to over 60), with differing cohorts (from 1971 to 2013) and in several countries (Australia, France, Estonia, Holland, UK, USA, Scandinavian countries, and South Korea). The representability of the analysed studies varied from convenience to random samples. In Finland, Flynn and Shayer used large-scale military test scores (Dutton & Lynn, 2013[1]), which showed gains before 2007 (average IQ gain 0.43 for years 1988–2007) and losses since 2007 (average IQ loss 0.25 per year). They presented the conclusions concerning Scandinavian countries as a case of the law of diminishing returns due to that

[1]The original publication has later been compromised as it reported unpublished results of another researcher without giving credit to the original author. This, however, does not alter the changes in the test scores during the time period.

> (i)n Scandinavia, the factors that have caused IQ gains may have exhausted their potency. Their educational system is more advanced and may show that at a certain point, schooling has reached a limit in terms of producing more graduates that can generalize and use logic on the hypothetical (mental abilities that pay dividends on IQ tests – Flynn, 2009). Further, good schooling has reached all classes and a more developed welfare state blunts the edge of class. Creation of cognitively demanding administrative jobs may have reached a limit beyond which economic efficiency forbids more "feather bedding". (Flynn & Shayer, 2018, p. 114)

Learning to Learn, Cognitive Abilities and Intelligence

In this study, we have adopted the concept and methods for studying anti-Flynn effect from intelligence studies and apply it to L2L research. L2L is the interplay of domain-general and domain-specific cognitive abilities, willingness to start to work and commitment to intellectually demanding tasks. The major difference to psychometrically oriented intelligence research is the focus on the relative malleability of cognitive abilities through educational means (c.f., Adey et al., 2007) and the contribution of motivational and situational factors to what we observe as performance scores in a certain task. In their comparative psychological paper "The Evolution of General Intelligence", Burkart and colleagues (2017, Figure 3) present a link that can be applied to our context: "The origin of cognitive skills … Empirical interrelations between brain size, general intelligence, and executive functions. The latter two entities are only visible to selection to the extent that they are translated into fitness-enhancing cognitive skills". Intelligence cannot be measured as a direct observation, neither can L2L, so the translation into *fitness-enhancing cognitive skills* can be redescribed as investments. The investment theory has been originally proposed by Cattell (1987), and further elaborated in educational studies by Kvist and Gustafsson (2008) and Demetriou and Spanoudis (2018), Demetriou et al. (2018). Measuring investments is complicated yet possible: solving a task calls for investment of abilities given already at birth (Plomin & von Stumm, 2018), cultivated knowledge and skills accumulated during educative processes, and motivation to accept the task and to do ones' best. Thus, the link between L2L and intelligence is sufficient to justify the application of the anti-Flynn concept. At the same time, the non-intelligence-related dimensions of L2L can provide new insights to understanding the anti-Flynn observations from intelligence studies as the investment approach allows one to speculate the non-cognitive reasons for them. Are they due to "material deteriorisation" (of intelligence, cognitive skills, or brain) or changed meanings and believes attached to measurement occasions? Are pupils and students willing to show the degree of adaptations (fitness-enhancements) of their accumulated resources, results of their own and others' investmental or educative activities? Every occasion, in which any intelligence test is used is always tapping also the approach/avoidance motivation and the meanings attached to the situation. Thus, cognitive tests can be safely said to measure only the lower limit of "intelligence" or any competence.

When previous anti-Flynn research is reviewed from this perspective, there is a comprehensive lack of studies in which the type of assessment situation (low- vs. high-stakes) is taken into account. Moreover, to our knowledge, the role of and changes in incentives, attitudes and beliefs on the fluctuation of cognitive scores have not been studied. Beliefs and attitudes are not only reflecting earlier experiences, they also include components of anticipated situations, which are also related to social and economic prospects. It is assumed in this paper that one of the reasons for downward trends in cognitive scores is a kind of anti-Achievement Motivation Effect, against the observed trends in line of McClelland and Atkinson's Theory of Achievement Motivation in advanced (pre)-industrial societies (McClelland et al., 1976).

Research Questions and Hypotheses

To provide plausible new explanations, we need to understand the effects of attitudes or beliefs, gender, and socio-economic, school level and geographical factors on performance. The present study is the first attempt to explain the decline in the Finnish context. In this study, we were focusing on the following hypotheses:

RQ1: What is the trend in Finnish students' performance in cognitive learning to learn assessment tasks based on nationally representative data from 2001, 2012 and 2017?

H1: We expect to find an anti-Flynn effect. Based on earlier Finnish studies, we expect to see the biggest decline between 2001 and 2012.

RQ2: Assuming that the anti-Flynn effect in cognitive performance is confirmed, do we see similar trends in learning-related beliefs?

H2: Students' mastery beliefs have declined and detrimental beliefs increased at the same rate as changes in cognitive performance.

RQ3: Can the changes in cognitive performance explain the changes in learning-related beliefs? Has their structure changed?

H3: Cognitive performance has a strong positive relation with mastery beliefs and a negative relation with detrimental beliefs. The changes in the structure largely explain the decline in performance.

Methods

Participants

In 2001, we took a random sample of Finnish lower secondary schools. Every ninth grader in the whole country had an equal probability to be included and when an individual student was selected, he/she drew all the other agemates in the school in the sample. The procedure was continued until we had over 10% of the whole student population ($N = 6279$ students in 82 schools) and had sufficient geographical coverage of the different parts of the country. All sampled schools agreed to participate. In 2012, 76 schools from the original sample still existed while the remaining six had been either merged to another school in the area or moved to a nearby location. We concluded that they were still close enough to the original sample schools and we included all ninth graders of those 82 schools in the 2012 sample ($N = 8875$). In 2017, we had $N = 9241$ students taking the test in the same schools. In 2001 and 2012, 43 randomly selected schools took the assessment entirely on a computer-based platform (CBA) while the remaining schools took the test either fully or partially on paper (PBA). In 2017, all students did the computer-based version of the test. In this study, we only report the data from the original CBA schools ($N = 2547$ / $N = 3729$ / $N = 3470$ in 2001, 2012 and 2017, respectively) as there was a considerable mode effect in 2001 (cognitive, but not affective PBA results were better than CBA). However, we ran all the analyses also with the full data, which did not change the interpretation of the main findings. The mean age of the students was 15.9 years (SD = .38) in each assessment. Gender distribution was equal each year. The proportion of Swedish-speaking students corresponded to the national population (approximately 5%). Approximately 5–10% of the originally sampled students were absent from school on the day of the assessment due to illness. As none of the sampled students or schools refused to participate, the missing data were random and they should not influence the interpretations of the results.

The assessments were conducted on assignment from the (Finnish) National Board of Education (2001) and the Ministry of Education and Culture (2012 and 2017), whose members also reviewed the quality and the ethics of the procedure and the measures. First, the sampled schools were contacted by a letter explaining the purpose and the procedure of the assessment. School principals appointed teachers to implement the assessment in each ninth grade class in the school and a material package containing detailed instructions, a letter for parents, and usernames/password for the assessment platform were sent to them. Teachers administered the test in school computer labs. Technical support was available upon request.

Measures

Cognitive Performance Tasks

The Finnish L2L test battery that has been used since the mid-1990s in both national languages (Finnish and Swedish) measures a wide scope of domain- general competences needed in all school subjects. We had 40 anchor items in four task types (deductive reasoning, evaluation of relevance of given information, quantitative reasoning, reading comprehension) in all assessment cycles (see Appendix A for descriptive statistics). There were two verbal reasoning tasks from the Ross Test of Higher Cognitive Processes (Ross & Ross, 1979). In the deductive reasoning task, the students were given two premises and they had to select the correct conclusion from among five answering options. The reliability of the six-item task was $\alpha = .51$, $\alpha = .60$ and $\alpha = .57$ in 2001, 2012 and 2017. In the second task (labelled Relevance in Appendix A), the students were given a verbal mathematical problem and some information for solving it. The task was not to solve the problem, just to evaluate if it was solvable based on the information they had been given. They also had to select whether there was enough, too little or even too much or irrelevant information. The ten items were scored as correct or incorrect ($\alpha = .52$, $\alpha = .51$ and $\alpha = .53$ in 2001, 2012 and 2017).

Quantitative reasoning was measured with seven items adapted from Sternberg's Triarchic abilities test (Sternberg et al., 2001). The students were presented with two conditionally defined invented mathematical concepts, *lag* and *sev* (e.g., if $x < y$, lag = +, otherwise lag = *), which they had to apply to the mathematical problems (i.e., 3 lag 2 = ?) to be solved. Each correctly solved item yielded a score. The reliability of the task was acceptable ($\alpha = .60$, $\alpha = .62$ and $\alpha = .64$ in 2001, 2012 and 2017).

The last anchor task assessed reading comprehension with 16 items based on the Kintsch and van Dijk (1978) macroprocessing framework (Lehto et al., 2001; Lyytinen & Lehto, 1998). The students read an expository text followed by items that were either good descriptions of the text as a whole, relevant pieces of information or irrelevant details. The students' task was to categorise the statements in these three groups. All the items were coded dichotomously as correct or incorrect ($\alpha = .67$, $\alpha = .66$ and $\alpha = .64$ in 2001, 2012 and 2017).

Learning-Related Beliefs

We had two scales for mastery beliefs (beliefs that are positively associated with learning) and three scales for detrimental beliefs (beliefs that are negatively associated with learning) that were identical in all assessment cycles. Descriptive statistics for individual items are presented in Appendix B. The first mastery beliefs scale was based on achievement goal theory (achievement orientation; Harackiewicz et al., 2002) and the second measured students' attitudes towards school. The measure for detrimental beliefs consisted of scales for avoidance orientation, self-handicapping (Urdan & Midgley, 2001) and causal beliefs about the role of luck in explaining success (Little et al., 2001). All scales consisted of three self-report items measured on a 7-point Likert scale. The items and the reliabilities of the scales are presented in Table 1.

Statistical Methods

We used SPSS24 for descriptive statistics and Mplus 7.2. (Muthén & Muthén, 2012) for all other analyses. First, we created item parcels using a homogeneous parcelling method (see Marsh et al., 2013) to reduce the number of parameters to be estimated. We averaged the individual items of a single test/scale and used the average score as a factor indicator in confirmatory factor analysis (CFA). Our criteria for an acceptable model fit was *CFI* and *TLI* > .95 and *RMSEA* < .06. Descriptive statistics for the item parcels are presented in Table 2.

In CFA, we had three factors, one for cognitive performance (with four indicators), and one each for mastery and detrimental beliefs. We used multiple group CFA treating the different data collection cycles as groups. We defined the 2001 cycle as the baseline level against which the other results

Table 1. Items in and reliabilities of the beliefs scales.

Achievement orientation ($\alpha = .87$, $\alpha = .89$ and $\alpha = .90$ in 2001, 2012 and 2017)
My goal is to do well at school.
Getting good marks at school is important to me.
For me, an important goal is to do well at school.
Attitudes towards school ($\alpha = .78$, $\alpha = .83$ and $\alpha = .85$ in 2001, 2012 and 2017)
I think we learn useful and important things at school.
I think we are taught a lot of interesting things at school.
I think our school is an efficient place for learning.
Avoidance orientation ($\alpha = .78$, $\alpha = .81$ and $\alpha = .80$ in 2001, 2012 and 2017)
I try to finish my homework with as little work as possible.
I only do the compulsory work for school, nothing more.
I have no interest in doing anything extra for school.
Self-handicapping ($\alpha = .71$, $\alpha = .75$ and $\alpha = .75$ in 2001, 2012 and 2017)
I get nervous when faced with a difficult problem.
Concentrating on difficult tasks is hard for me.
I give up easily if my assignments look too demanding.
Causal beliefs (luck) ($\alpha = .66$, $\alpha = .79$ and $\alpha = .75$ in 2001, 2012 and 2017)
Success at school is a matter of luck.
One cannot really influence one's success at school.
Failure at school is mainly due to bad luck.

were compared. Before comparing latent means, we tested measurement invariance across data collection cycles by first constraining factor loadings and then intercepts equal and studying changes in fit indices. We tested the first and the second hypothesis by comparing latent means of the constrained models. For testing the third hypothesis, we fitted a multiple group structural equation model on the data. In all analyses, we produced confidence intervals for the estimates by bootstrapping with 1000 replicates (Cheung & Lau, 2008).

Results

Measurement Invariance Testing

We specified a model with three factors: cognitive performance (four indicators), mastery beliefs (two indicators) and detrimental beliefs (three indicators). Due to the small number of indicators per factor, we had to test the measurement invariance of all three factors in one analysis by letting the factors correlate with each other. The first attempt to fit the baseline model on the data yielded a relatively poor model fit ($CFI = .924$; $TLI = .886$; $RMSEA = .076$; $\chi^2 = 1434.668$, df = 72, $p < .001$), but allowing a residual correlation of Avoidance orientation and Cognitive performance improved the model fit considerably. The further fit indices of the stepwise measurement invariance testing are displayed in Table 3. The table shows that constraining factor loadings equal across groups even improved TLI and RMSEA, but we had to release two constraints on the intercepts of the cognitive data from 2001 to get an acceptable model fit. In practice, this means that while all the other intercepts were approximately equal to the means of 2001 as displayed in Table 2, the intercept for quantitative reasoning (Math) was higher and the intercept for reading comprehension was lower in 2012 and 2017 compared to the baseline of 2001. Thus, in 2001 the students' performance on the Math items was not as high as the latent mean differences would suggest; on the other hand, students in the 2001 sample performed even better than expected on the reading comprehension task. We used this almost fully constrained model in all further analyses, but when interpreting the results, we have to keep in mind the implications of the two released constraints (cf., Härnqvist et al., 1994).

Trends in Cognitive Performance

Our first hypothesis was that we would see a declining trend in students' performance in the cognitive tasks compared to the baseline level of 2001. We tested the hypothesis by studying changes in

Table 2. Descriptive statistics for the item parcels.

	2001					2012					2017				
Item parcel	*N*	Min	Max	*M*	SD	*N*	Min	Max	*M*	SD	*N*	Min	Max	*M*	SD
Deductive	2509	0	100	56.46	25.21	3726	0	100	49.63	27.89	3459	0	100	52.09	26.76
Relevance	2496	0	100	55.47	19.31	3714	0	100	51.59	19.33	3442	0	100	51.31	19.73
Math	2468	0	100	45.34	25.94	3687	0	100	47.85	26.41	3339	0	100	48.64	27.10
Reading	2441	0	100	54.20	19.83	3643	0	100	47.46	19.72	3272	0	100	48.38	19.36
Achievement orientation	2499	1	7	5.52	1.15	3710	1	7	5.20	1.23	3433	1	7	5.34	1.25
Attitudes towards school	2523	1	7	4.76	1.09	3729	1	7	4.50	1.18	3480	1	7	4.70	1.19
Self-handicapping	2499	1	7	3.70	1.26	3710	1	7	3.94	1.27	3435	1	7	3.97	1.35
Avoidance orientation	2500	1	7	4.18	1.37	3709	1	7	4.45	1.32	3434	1	7	4.20	1.38
Causal beliefs (luck)	2500	1	7	2.22	1.08	3710	1	7	2.41	1.26	3435	1	7	2.29	1.19

N = number of students, Min = smallest value, Max = largest value, *M* = mean, SD = standard deviation.

Table 3. Measurement invariance of the model.

Factor	CFI	TLI	RMSEA	χ^2	df	*p*
Baseline model	.964	.944	.053	708.555	69	<.001
Factor loadings constrained equal	.961	.952	.049	781.804	87	<.001
Intercepts constrained equal	.938	.933	.059	1208.266	99	<.001
Two cognitive intercepts released for the 2001 data	.956	.951	.050	882.387	97	<.001

Table 4. Latent means and their confidence intervals compared to the baseline level (0) of 2001.

	Cognitive tasks			Mastery beliefs			Detrimental beliefs		
Year	Mean	CI low	CI high	Mean	CI low	CI high	Mean	CI low	CI high
2012	−.333	−.395	−.273	−.336	−.400	−.273	.335	.263	.402
2017	−.243	−.305	−.177	−.144	−.208	−.082	.167	.094	.234

CI = 95% confidence interval.

latent means in the model with three correlating factors (cognitive performance, mastery beliefs, detrimental beliefs). We tested the statistical significance of the differences by bootstrapping 95 percent confidence intervals for all estimates. The latent means and their confidence intervals for all scales are displayed in Table 4 and Figure 1.

Table 4 and Figure 1 show that students' performance declined by −.333 factor points between 2001 and 2012. Confidence intervals of the estimate did not overlap with the baseline, so the change was statistically significant. In 2017, students' performance had slightly increased again, but the difference between it and the 2001 baseline was still statistically significant. The increase between 2012 and 2017 was quite small, but the estimates did not fall between each other's confidence intervals. However, the higher confidence interval of 2012 and lower confidence interval of 2017 overlapped, indicating that the increase was not large. Nevertheless, the results confirmed our hypothesis about the strong decline in performance between 2001 and 2012. Furthermore, they confirmed our assumption that the trend had reversed since 2012.

Trends in Learning-Related Beliefs

Our second hypothesis was that if we confirmed the anti-Flynn effect in cognitive performance, we would also observe similar trends in learning-related beliefs. Means for mastery and detrimental beliefs are presented together with their 95 percent confidence intervals in Table 5 and Figure 2.

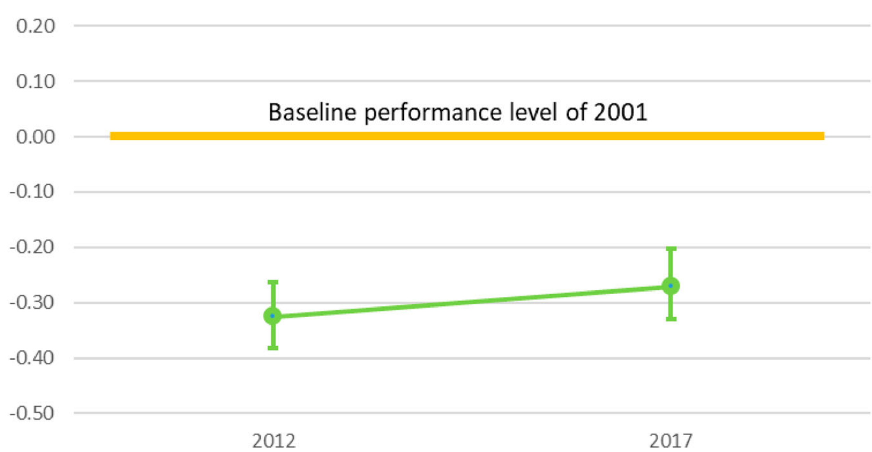

Figure 1. Trends of performance in cognitive tasks. Error bars represent 95% confidence intervals.

Table 5. Standardised and unstandardised estimates and their 95% confidence intervals.

	Cognitive on mastery				Cognitive on detrimental				Mastery with detrimental		
Year	*β*	*B*	CI low	CI high	*β*	*B*	CI low	CI high	*r*	CI low	CI high
2001	−.043 ns	−.060	−.275	.123	−.725***	−1.008	−1.231	−.825	−.703	−.776	−.622
2012	.228***	.341	.212	.453	−.602***	−.901	−1.060	−.765	−.513	−.584	−.442
2017	.225***	.343	.227	.452	−.615***	−.938	−1.069	−.830	−.515	−.584	−.436

r = correlation; CI = 95% confidence interval, ns = non-significant, ****p* < .001.

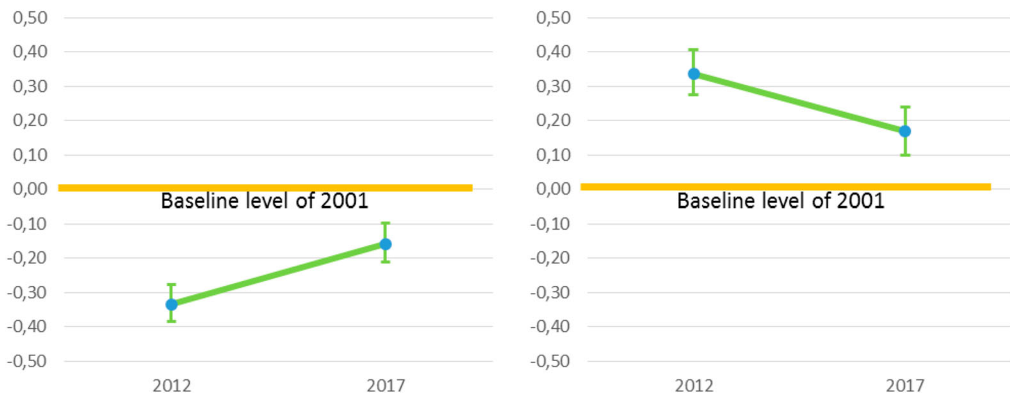

Figure 2. Trends in learning-related mastery (left figure) and detrimental (right figure) beliefs. Error bars represent 95% confidence intervals.

They show the expected pattern: mastery beliefs declined and detrimental beliefs increased from 2001 to 2012, but after that the trend reversed again. In 2017, mastery beliefs were statistically significantly higher and detrimental beliefs lower than in 2012, but they were still below/above (respectively) the baseline level of 2001. Thus, our second hypothesis was also confirmed.

Structure of Cognitive Performance and Learning-Related Beliefs

Our third hypothesis was that cognitive performance has a strong positive relation with mastery beliefs and a negative relation with detrimental beliefs. We also assumed that the changes that happened in the *structuration* of them would explain why performance has declined. To test this hypothesis, we specified a structural equation model in which we predicted the cognitive performance factor on mastery and detrimental beliefs that were allowed to correlate with each other. To test the significance of differences in coefficients, we bootstrapped 95 percent confidence intervals on the estimates. We used the same model modifications as above. The model fit the data well (*CFI* = .957, *TLI* = .952, *RMSEA* = .050, χ^2 = 873.494, df = 97, $p < .001$). The model explained 48.3, 55.4 and 57.1 percent of the variance of cognitive performance in 2001, 2012 and 2017, respectively.

The results of the model are presented in Table 5. We see a statistically significant change in structuration of learning-related beliefs and performance between 2001 and 2017. In 2001, mastery and detrimental beliefs correlated negatively and strongly, and detrimental beliefs were a strong negative predictor of cognitive performance. Mastery beliefs did not have a unique contribution to explaining performance when detrimental beliefs were taken into account. In 2012, the negative correlation between mastery and detrimental beliefs was still relatively strong but statistically significantly weaker than in 2001. This means that there were more controversial students who had both mastery and detrimental beliefs. In 2012, mastery beliefs had become a statistically significant predictor of performance while detrimental beliefs still predicted it negatively too. Although the negative effect of detrimental beliefs on performance looked slightly weaker than in 2001, confidence interval comparison showed that this change was not statistically significant. The 2017 results were almost identical to those of 2012, so the structure had remained the same between the two latter time points. The changes observed between 2001 and 2012 confirmed our third hypothesis about the influence of changing beliefs structure on students' performance in the assessment situation.

Discussion

The aim of this study was to confirm earlier observations about the declining trend in Finnish students' performance in educational assessments, to understand *when* the decline has taken place, and to find explanations for the decline in the changing beliefs structures of the students. Therefore, we tested three hypotheses using multiple group structural equation modelling on three waves of nationally representative data from computer-based learning to learn assessments of 15-year-old ninth grade students in 2001, 2012 and 2017.

In our first hypothesis, we expected to find an anti-Flynn effect. As we had previous non-representative municipal assessment results available (Marjanen et al., 2014), we expected to see the biggest decline between the first two assessment cycles. The results confirmed this hypothesis: students' performance in the cognitive L2L assessment task declined by .33 factor points between 2001 and 2012, whereas between the two latter time points performance was slightly increasing again. In the municipal assessment that served as an inspiration for this study, the decline could be dated between 2004 and 2010 (Marjanen et al., 2014) and also the trends in the Finnish PISA results show a similar pattern (e.g., Leino et al., 2019). In PISA, the Finnish performance scores were increasing until 2006, but after that, the trend has turned downwards. Until the 2018 cycle, PISA has not yet shown a new upward trend, but the results of our present study suggest that this may be possible.

But what makes the results fluctuate like this? A "true" anti-Flynn effect would imply that the cognitive competences of students have deteriorated for some reason, but the explanation does not quite fit with the finding of a reversing trend. Additionally, previous international studies about anti-Flynn effects discuss the possibilities that there might be explanations not directly related to cognitive competences behind this phenomenon (Flynn & Shayer, 2018). Therefore, we continued by testing further hypotheses about the changing beliefs structures of Finnish students as a potential explanation for the decline. There is a lot of evidence that learning-related beliefs are associated with student' performance in educational assessments (Eccles & Wigfield, 2002). Students with high levels of *mastery beliefs* perform better than students with a high level of *detrimental beliefs* (Harackiewicz et al., 2002), but these two sets of beliefs do not necessarily form a continuum as there may be students who have high goals and aspirations but who do not manage to put in enough effort to reach these goals. First, we tested the hypothesis that there might be changes in the overall level of mastery and detrimental beliefs between the three assessment cycles. After that, our last hypothesis was that the structural changes in these beliefs might be the key to understanding the fluctuations in cognitive performance.

Testing the second hypothesis clearly showed that the changes in the trends of mastery and detrimental beliefs followed the pattern demonstrated by cognitive performance. Declining performance was associated with a lower level of mastery attitudes and higher level of detrimental beliefs and the slight turn of the trend back towards a more positive situation was reflected in both belief dimensions. This finding is aligned with earlier cross-sectional studies about the relations of beliefs and cognitive test scores (see Kupiainen et al., 2014, for a review). However, to our knowledge this is the first study to show how the *structuration* of mastery and detrimental beliefs has changed over a longer time span with repeated cross-sectional data sets from the same schools and how these changes explain the changing cognitive trends. The importance of mastery beliefs in predicting cognitive performance has increased. From another perspective, this shows that in 2001 even students with lower levels of mastery beliefs were doing well on the test while this was no longer the situation in 2012. Detrimental beliefs were a strong negative predictor of cognitive performance at all time points. The weakening in the correlation between mastery and detrimental beliefs indicates that in 2012 and 2017 there were more controversial students who had both mastery and detrimental beliefs compared to the baseline of 2001 when the negative correlation between these constructs was very strong. In other words, in the two latter measurements there were more students who wanted to do well but were nevertheless not willing to put the necessary effort into the test and their schoolwork in general (cf., Carroll, 1963; Kupiainen et al., 2014; see also Wise & DeMars, 2010). This interpretation was supported by the finding that an improvement in beliefs in 2017 was associated with increasing cognitive scores even though this difference was not quite statistically significant.

Our study had some limitations that need to be taken into account when interpreting the validity of our conclusions. The biggest limitation is related to the measurement invariance being reached only partially as we had to release the intercepts of math and reading scores from 2001 to conduct the further analyses. This was due to the "too high" reading and "too low" math performance of the 2001 sample in relation to the overall CFA model of the whole data. This means that if the different cognitive domains were analysed separately the effects could have been even stronger for reading, whereas for math the findings presented here do not necessarily apply on a full scale. However, there is recent evidence from studies using the same measures as in the present study that students' math performance scores are also influenced by beliefs and differential effort in the assessment situation (Vainikainen & Hautamäki, 2018). Therefore, a follow-up study should look into the different areas of cognitive performance in more detail to create a more fine-tuned understanding of the mechanisms underlying effects of learning-related beliefs on cognitive performance in a low-stakes assessment situation.

Another limitation of the study is that in 2001 and 2012 some schools took the tests on paper (PBA) and others on a computer-based platform (CBA). In 2001, the assessment technology was less developed and a strong mode effect made the comparison of PBA and CBA results impossible. Therefore, we could only use the data from CBA schools even if at later time points we would have had much more comparable data from original PBA schools. Indeed, we also ran the analyses with the full data, ignoring the 2001 mode effects, but this did not change the key conclusions about the anti-Flynn effect. The findings about the structuration remained almost exactly the same, but the decline of cognitive performance was even more dramatic due to the very high test results of the 2001 PBA schools. In the future, some correction methods for the mode effect could be applied to utilise the whole datasets instead of just a subsample of them.

Conclusions

Our study was the first representative large-scale study showing that changes in the structures of students' learning-related beliefs are associated with fluctuating performance levels in a low-stakes cognitive testing situation. The results also confirmed our hypothesis that the overall decline in performance and beliefs during the first decade of the 2000s was not a local phenomenon in one southern Finnish well-off municipality, but it applied to the whole country. The results of this and repeated local studies show that the decline has now stopped and there is a slight trend towards a more positive situation again. As beliefs and performance follow the same trends and there is a relatively strong relation between them, our interpretation is that we may not be dealing with a "true" anti-Flynn effect. Instead, it seems that Finnish students' approach to schoolwork and effort in assessment situations has fluctuated and this could explain the changes in demonstrated performance in national and international assessment studies.

To outline some consequences and to search for remedies to the repeatedly observed downward trends of cognitive scores, there is an option to reopen the issue of relations between education and psychology. Very generally, this situation could be described as a return to Cronbach and Snow's Aptitude-Treatment Interaction (ATI) theory (1977). The present version might be called *a* *developmentally sensitive ATI-model for age-graded education from kindergarten to higher education*. The most extensive presentation about this is given by Demetriou and his team (e.g., 2018a; 2018b). In this conceptualisation, L2L would be a significant interaction effect.

The results of this study can be applied in four different ways. First, we advocate the developmental approach to education through analysing the developmental, time/age-related relations of self-beliefs, cognition and school achievement. These should be considered especially when transitions from primary to secondary and from secondary to tertiary education are treated as "prepared fitness" to benefit from introduction to the more advanced contents of disciplines. This means also the need to reanalyse the match of developmental level of students to the cognitive demands (loads) of school disciplines (Shayer & Adey, 1981).

Secondly, it is important to take seriously the need for teaching for thinking: different disciplines have partially different dominant ways and forms of thinking (explanation vs understanding; historical, mathematical, logical, experimental, social and juridical as well artistic thinking etc.). The problem of the specific and general in thinking is the core of this second aspect (Adey et al., 2007;, Greiff et al., 2014; Vainikainen et al., 2015). Klauer (2000) refers to this as Huckeback theorem, concluding that one can teach specifics with generic transfer-effects, but not the opposite. Thirdly, Lewin (1935) pointed out already in 1935 in his dynamic personality theory that school examinations, gradings and diplomas are objects which have both positive and negative valance; the hope of success and the fear of failure. For this reason, we have used several scales for measuring self-concepts/-beliefs to take into account the personal interpretations of meanings of achievement situations and tasks. Therefore we also define L2L as a two-vector mental frame, the mastery of thinking and the perspective of hope, to accept a task and adapt the present competencies to the task at hand. L2L is not only a large set of self-beliefs, not only a cognition, but both: competence and willingness, as we define it, to adapt to the situation, a match of the internal self and outside world, the generalised aptitude-treatment-interaction.

Finally, we propose that the L2L tools can provide an added-value in relation to school marks – and other assessments like PISA – in predicting the survival capital (Rindermann, 2018) of individuals well-educated in their respective nations, to represent a general estimate of nations' human capital. We agree with Demetriou (in press) when he concludes that changes in principle-based and critical thought (in primary and secondary education, our addition) are needed by many (more, our addition) students if they (the nation, we add) would be able to grasp science as intended by universities (and scholastic disciplines). According to Demetriou and colleagues (2020, p. 1)

> learning in different domains, such as language and mathematics, depends on an interaction between the general cognitive processes dominating in each cycle and the state of the symbol systems associated with this domain. If command of any of these systems is deficient, specific learning deficiencies may emerge, as in dyslexia and dyscalculia.

Learning-to-learn is not to be reduced to a new scholastic discipline, but it is to be understood as the added-value of well-designed education, as the statistically and realistically significant interaction term in generalised aptitude-treatment-interaction framework for schooling. It has true predictive power as such, complementing the school achievement. Therefore, the downward trends in PISA and other measures of key competencies pose a threat to the prospects of prosperity of Finland and all other countries, in which anti-Flynn effects have been observed. This is the impetus to study this topic, also in coming studies.

Acknowledgements

We want to thank the whole staff of the Centre for Educational Assessment at the University of Helsinki for the practical implementation of the three studies reported in this article.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Mari-Pauliina Vainikainen http://orcid.org/0000-0003-4186-692X

References

Adey, P., Csapó, B., Demetriou, A., Hautamäki, J., & Shayer, M. (2007). Can we be intelligent about intelligence? Why education needs the concept of plastic general ability. *Educational Research Review*, *2*, 75–97. https://doi.org/10.1016/j.edurev.2007.05.001

Atkinson, J. W. (1964). *An introduction to motivation*. D. van Nostrand Company, Inc.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.

Baraldi, C., & Corsi, G. (2017). *Niklas Luhmann: Education as a social system*. Springer Briefs on Key Thinkers in Education. https://doi.org/10.1007/978-3-319-49975-8_5

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017). The evolution of general intelligence. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/S0140525X16000959

Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, *64*, 723–733.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Elsevier.

Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables. Bootstrapping with structural equation models. *Organizational Research Methods*, *11*(2), 296–325. https://doi.org/10.1177/1094428107300343

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington Publishers, Inc.

Demetriou, A., Greiff, S., Makris, N., Spanoudis, G., Panaoura, R., & Kazi, S. (2020). Bridging educational priorities with developmental priorities: Towards a developmental theory of instruction. https://doi.org/10.31234/osf.io/kjmu3.

Demetriou, A. (in press). Bridging the 21st century gap in education: History, causation, and solutions. *European Review*. https://doi.org/10.31234/osf.io/q5mrj

Demetriou, A., Makris, N., Kazi, S., Spanoudis, G., Shayer, M., & Kazali, E. (2018). Mapping the dimensions of general intelligence: An integrated differential-developmental theory. *Human Development*, *61*, 4–42. https://doi.org/10.1159%2F000484450

Demetriou, A., & Spanoudis, G. (2018). *Growing minds: A general theory of intelligence and learning*. Routledge.

Dutton, E., & Lynn, R. (2013). A negative Flynn effect in Finland, 1997–2009. *Intelligence*, *41*, 817–820. https://doi.org/10.1016/j.intell.20136.05.008

Eccles, J., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Flynn, J. R. (1984). The mean of IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*(1), 29–51. https://doi.org/10.1037/0033-2909.95.1.29

Flynn, J. R., & Shayer, M. (2018). IQ decline and Piaget: Does the rot start at the top? *Intelligence*, *66*, 112–121. https://doi.org/10.1016/j.intell.2017.11.010

Fredriksson, U., & Hoskins, B. (2007). The development of learning to learn in a European context. *The Curriculum Journal*, *18*(2), 127–134. https://doi.org/10.1080/09585170701445921

Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, H., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review*, *13*, 74–83. https://doi.org/10.1016/j.edurev.2014.10.002

Greiff, S., Wüstenberg, S., Goetz, T., Vainikainen, M.-P., Hautamäki, J., & Bornstein, M. (2015). A longitudinal study of higher-order thinking skills: Working memory and fluid reasoning in childhood enhance complex problem solving in adolescence. *Frontiers in Psychology*, *6*, 1060. https://doi.org/10.3389/fpsyg.2015.01060

Griffin, P., & Care, E. (2015). *Assessment and teaching of 21st century skills: Methods and approach*. Springer. https://doi.org/10.1007/978-94-017-9395-7

Habók, A. (2015). Learning to learn in years 1 and 2 of Hungarian primary schools. *Education 3-13*, *43*(2), 153–163. https://doi.org/10.1080/03004279.2013.783875

Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology*, *94*(3), 638–645. https://doi.org/10.1037/0022-0663.94.3.638

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, *56*(1), 51–56. https://doi.org/10.1037/h0062474

Härnqvist, K., Gustafsson, J.-E., Muthén, B. O., & Nelson, G. (1994). Hierarchical models of ability at individual and class levels. *Intelligence*, *18*, 165–187. https://doi.org/10.1016/0160-2896(94)90026-4

Hautamäki, A., Hautamäki, J., & Kupiainen, S. (2010). Assessment in schools - learning to learn. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (pp. 268–272). Elsevier Scientific Publ. Co.

Hautamäki, J., Arinen, P., Eronen, S., Hautamäki, A., Kupiainen, S., Lindblom, B., & Scheinin, P. (2002). *Assessing learning-to-learn: A framework*. National Board of Education, Evaluation 4/2002.

Hautamäki, J., Kupiainen, S., Marjanen, J., Vainikainen, M.-P., & Hotulainen, R. (2013). *). Oppimaan oppiminen peruskoulun päättövaiheessa: Tilanne vuonna 2012 ja muutos vuodesta 2001 [Learning to learn at the end of basic education: Situation in 2012 and change from 2001]*. University of Helsinki. Department of Teacher Education Research Report 347. Unigrafia.

Hoskins, B., & Deakin Crick, R. (2010). Competences for learning to learn and active citizenship: Different currencies or two sides of the same coin? *European Journal of Education*, *45*(1), 121–137.

Hullemann, C. S., Durik, A. M., Schweigert, S. A., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology*, *100*(2), 398–416. https://doi.org/10.1037/0022-0663.100.2.398

Kintsch, W., & van Dijk, T. A. (1978). Towards a model of text comprehension and production. *Psychological Review*, *8*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Klauer, K. J. (2000). Das Huckepack-Theorem asymmetrischen Strategietransfers [The piggyback theorem of asymmetric strategy transfer. A contribution to the theory of training and transfer]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *32*(3), 153–165. https://doi.org/10.1026/0049-8637.32.3.153

Kupiainen, S., Marjanen, J., Vainikainen, M.-P., & Hautamäki, J. (2011). *Oppimaan oppiminen Vantaan peruskouluissa: Kolmas-, kuudes- ja yhdeksäsluokkalaiset oppijoina keväällä 2010 [Learning to learn in Vantaa: The third, sixth and ninth graders as learners in Spring 2010]*. Vantaan kaupungin Sivistysvirasto [Education Department of the City of Vantaa].

Kupiainen, S., Vainikainen, M.-P., Marjanen, J., & Hautamäki, J. (2014). The role of time on task in computer-based low-stakes assessment of cross-curricular skills. *Journal of Educational Psychology*, *106*(3), 627–638. https://doi.org/10.1037/a0035507

Kvist, A. V., & Gustafsson, J. E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence*, *36*(5), 422–436.

Lai, E. R., & Viering, M. (2012). *Assessing 21st century skills: Integrating research findings*. National Council on Measurement in Education.

Lehto, J., Scheinin, P., Kupiainen, S., & Hautamäki, J. (2001). National survey of reading comprehension in Finland. *Journal of Research in Reading*, *24*(1), 99–110. https://doi.org/10.1111/1467-9817.00135

Leino, K., Ahonen, A., Hienonen, N., Hiltunen, J., Lintuvuori, M., Lähteinen, S., Lämsä, J., Nissinen, K., Nissinen, V., Puhakka, E., Pulkkinen, J., Rautopuro, J., Sirén, M., Vainikainen, M.-P., & Vettenranta, J. (2019). *PISA 2018 ensituloksia: Suomi parhaiden joukossa [PISA 2018 first results: Finland among the top performers]. Publications of the Ministry of Education and Culture 2019:40*. Ministry of Education and Culture.

Lewin, K. (1935). The psychological situation of reward and punishment. In K. Lewin (Ed.), *A dynamic theory of personality: Selected papers* (pp. 114–170). McGraw Hill.

Little, T. D., Lopez, D. F., Oettingen, G., & Baltes, P. B. (2001). A comparative-longitudinal study of action-control beliefs and school performance. *International Journal of Behavioural Development*, *25*(3), 237–245. https://doi.org/10.1080/01650250042000258

Lyytinen, S., & Lehto, J. E. (1998). Hierarchy rating as a measure of text macroprocessing: Relationship with working memory and school achievement. *Educational Psychology*, *18*(2), 157–169. https://doi.org/10.1080/0144341980180202

Marjanen, J., Vainikainen, M.-P., Kupiainen, S., Hotulainen, R., & Hautamäki, J. (2014). *Oppimaan oppiminen Vantaan peruskouluissa: Kolmas-, kuudes- ja yhdeksäsluokkalaiset oppijoina vuosina 2013 ja 2010. [Learning to learn in comprehensive schools in Vantaa: Third, sixth and ninth graders as learners in 2013 and 2010]*. Vantaan kaupungin Sivistysvirasto [Education Department of the City of Vantaa].

Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, *80*(3), 366–380. https://doi.org/10.1037/0022-0663.80.3.366

Marsh, H. W., Lüdtke, O., Nagenkast, B., Morin, A. J. S., & Von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right – Camouflaging misspecification with item parcels in CFA models. *Psychological Methods*, *18*(3), 257–284. https://doi.org/10.1037/a0032773

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1976). *The achievement motive*. Irvington.

Moreno, A., & Martín, E. (2007). The development of learning to learn in Spain. *The Curriculum Journal*, *18*(2), 175–193.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide*. Version 7.

National Board of Education. (1999). *A framework for evaluating educational outcomes in Finland*. National Board of Education, Evaluation 8/1999.
OECD. (2001). *Knowledge and skills for life: First results from the OECD programme for international student assessment (PISA) 2000*. OECD Publishing. https://doi.org/10.1787/9789264266490-en
OECD. (2016). *PISA 2015 results (Volume I): Excellence and equity in education*. OECD Publishing. https://doi.org/10.1787/9789264266490-en
Pask, G. (1976). Styles and strategies of learning. *British Journal of Educational Psychology*, *46*(2), 128–148. https://doi.org/10.1111/j.2044-8279.1976.tb02305.x
Pellegrino, J. W., & Hilton, M. L. (eds.). (2012). *Education for life and work: developing transferable knowledge and skills in the 21st century*. National Research Council. Committee on Defining Deeper Learning and 21st Century Skills, Board on Testing and Assessment and Board on Science Education, Division of Behavioral and Social Sciences and Education. The National Academies Press.
Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, *19*(3), 148. https://doi.org/10.1038/nrg.2017.104
Rindermann, H. (2018). *Cognitive capitalism. Human capital and the wellbeing of nations*. Cambridge University Press.
Ross, J. D., & Ross, C. M. (1979). *Ross test of higher cognitive processes*. Academic Therapy Publications.
Shayer, M., & Adey, P. (1981). *Towards a science of science teaching: Cognitive development and curriculum demand*. Heinemann.
Sternberg, R., Castejon, J. L., Prieto, M. D., Hautamäki, J., & Grigorenko, E. (2001). Confirmatory factor analysis of the Sternberg Triarchic abilities test in three international samples. *European Journal of Psychological Assessment*, *17*, 1–16. https://doi.org/10.1027//1015-5759.17.1.1
Urdan, T., & Midgley, M. (2001). Academic self-handicapping: What we know and what more there is to learn. *Educational Psychology Review*, *13*(2), 115–138. https://doi.org/10.1023/A:1009061303214
Vainikainen, M.-P., & Hautamäki, J. (2018). Selittääkö yrittäminen oppilaiden osaamisessa havaittuja ryhmäeroja? Itsearvioitu yrittäminen, investoitu työaika ja osaamiserot lokitietoanalyysin valossa. [Are group differences in perfomance explained by effort? Self-reported effort, time investment and performance differences in the light of log data analysis]. *Psykologia*, *53*(2-3), 152–165.
Vainikainen, M.-P., Hautamäki, J., Hotulainen, R., & Kupiainen, S. (2015). General and specific thinking skills and schooling: Preparing the mind to new learning. *Thinking Skills and Creativity*, *18*, 53–64. https://doi.org/10.1016/j.tsc.2015.04.006
Van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2017). The relation between 21st century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, *72*, 577–588. https://doi.org/10.1016./j.chb.2017.03.010
Visentin, M. (2017). Learning to learn. *Italian Journal of Sociology of Education*, *9*, 1. https://doi.org/10.14658/pupj-ijse-2017-1-11
Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, *44*(3), 299–321. https://doi.org/10.1080/00220272.2012.668938
Wise, S. L., & DeMars, C. E. (2010). Examinee noneffort and the validity of program assessment results. *Educational Assessment*, *15*, 27–41. https://doi.org/10.1080/10627191003673216

Appendices

Appendix A. Descriptive Statistics for the Cognitive Items

	2001					2012					2017				
Item	*N*	Min	Max	*M*	SD	*N*	Min	Max	*M*	SD	*N*	Min	Max	*M*	SD
Deductive 1	2508	0	1	0.45	0.50	3722	0	1	0.35	0.48	3443	0	1	0.37	0.48
Deductive 2	2507	0	1	0.85	0.36	3709	0	1	0.68	0.46	3434	0	1	0.76	0.43
Deductive 3	2509	0	1	0.54	0.50	3706	0	1	0.59	0.49	3432	0	1	0.61	0.49
Deductive 4	2509	0	1	0.48	0.50	3703	0	1	0.40	0.49	3431	0	1	0.40	0.49
Deductive 5	2508	0	1	0.71	0.46	3696	0	1	0.61	0.49	3427	0	1	0.67	0.47
Deductive 6	2507	0	1	0.36	0.48	3715	0	1	0.35	0.48	3440	0	1	0.34	0.48
Relevance 1	2484	0	1	0.50	0.50	3709	0	1	0.49	0.50	3401	0	1	0.48	0.50
Relevance 2	2492	0	1	0.83	0.38	3698	0	1	0.76	0.43	3408	0	1	0.74	0.44
Relevance 3	2489	0	1	0.88	0.33	3692	0	1	0.86	0.35	3402	0	1	0.85	0.36
Relevance 4	2493	0	1	0.61	0.49	3675	0	1	0.64	0.48	3406	0	1	0.65	0.48
Relevance 5	2484	0	1	0.62	0.49	3678	0	1	0.55	0.50	3418	0	1	0.53	0.50

(*Continued*)

Continued.

Item	2001 N	Min	Max	M	SD	2012 N	Min	Max	M	SD	2017 N	Min	Max	M	SD
Relevance 6	2490	0	1	0.22	0.41	3686	0	1	0.14	0.35	3413	0	1	0.11	0.32
Relevance 7	2493	0	1	0.18	0.39	3679	0	1	0.17	0.37	3420	0	1	0.17	0.37
Relevance 8	2492	0	1	0.73	0.44	3680	0	1	0.66	0.47	3412	0	1	0.64	0.48
Relevance 9	2492	0	1	0.55	0.50	3692	0	1	0.44	0.50	3402	0	1	0.44	0.50
Relevance 10	2493	0	1	0.45	0.50	3702	0	1	0.50	0.50	3416	0	1	0.57	0.50
Math 1	2458	0	1	0.42	0.49	3676	0	1	0.45	0.50	3331	0	1	0.49	0.50
Math 2	2441	0	1	0.19	0.39	3670	0	1	0.21	0.40	3312	0	1	0.21	0.41
Math 3	2413	0	1	0.69	0.46	3604	0	1	0.70	0.46	3237	0	1	0.72	0.45
Math 4	2430	0	1	0.56	0.50	3654	0	1	0.59	0.49	3275	0	1	0.57	0.50
Math 5	2442	0	1	0.47	0.50	3663	0	1	0.45	0.50	3272	0	1	0.42	0.49
Math 6	2437	0	1	0.51	0.50	3624	0	1	0.57	0.50	3255	0	1	0.60	0.49
Math 7	2454	0	1	0.35	0.48	3649	0	1	0.41	0.49	3270	0	1	0.43	0.50
Reading 1	2439	0	1	0.48	0.50	3633	0	1	0.53	0.50	3257	0	1	0.51	0.50
Reading 2	2433	0	1	0.70	0.46	3637	0	1	0.59	0.49	3262	0	1	0.61	0.49
Reading 3	2434	0	1	0.58	0.49	3633	0	1	0.51	0.50	3262	0	1	0.55	0.50
Reading 4	2435	0	1	0.47	0.50	3624	0	1	0.32	0.47	3261	0	1	0.32	0.47
Reading 5	2433	0	1	0.68	0.47	3633	0	1	0.59	0.49	3262	0	1	0.62	0.49
Reading 6	2430	0	1	0.42	0.49	3624	0	1	0.38	0.49	3256	0	1	0.39	0.49
Reading 7	2434	0	1	0.45	0.50	3628	0	1	0.31	0.46	3259	0	1	0.35	0.48
Reading 8	2430	0	1	0.50	0.50	3631	0	1	0.40	0.49	3257	0	1	0.39	0.49
Reading 9	2437	0	1	0.72	0.45	3631	0	1	0.61	0.49	3252	0	1	0.62	0.49
Reading 10	2434	0	1	0.65	0.48	3633	0	1	0.64	0.48	3254	0	1	0.65	0.48
Reading 11	2432	0	1	0.47	0.50	3632	0	1	0.47	0.50	3255	0	1	0.47	0.50
Reading 12	2426	0	1	0.39	0.49	3632	0	1	0.34	0.47	3254	0	1	0.32	0.47
Reading 13	2427	0	1	0.34	0.47	3631	0	1	0.25	0.43	3253	0	1	0.27	0.44
Reading 14	2428	0	1	0.70	0.46	3634	0	1	0.56	0.50	3252	0	1	0.58	0.49
Reading 15	2429	0	1	0.48	0.50	3629	0	1	0.52	0.50	3249	0	1	0.53	0.50
Reading 16	2429	0	1	0.67	0.47	3632	0	1	0.58	0.49	3243	0	1	0.60	0.49

Appendix B. Descriptive Statistics for the Items in the Beliefs Scales

Item	2001 N	Min	Max	M	SD	2012 N	Min	Max	M	SD	2017 N	Min	Max	M	SD
Achievement orientation															
Item 1	2490	1	7	5.81	1.22	3704	1	7	5.39	1.31	3426	1	7	5.50	1.30
Item 2	2495	1	7	5.38	1.32	3705	1	7	5.17	1.34	3418	1	7	5.36	1.36
Item 3	2489	1	7	5.39	1.34	3697	1	7	5.03	1.42	3418	1	7	5.17	1.43
Attitudes towards school															
Item 1	2514	1	7	5.13	1.22	3728	1	7	4.63	1.36	3470	1	7	4.83	1.35
Item 2	2516	1	7	4.42	1.34	3726	1	7	4.36	1.37	3462	1	7	4.52	1.35
Item 3	2512	1	7	4.72	1.35	3729	1	7	4.50	1.37	3462	1	7	4.75	1.33
Avoidance orientation															
Item 1	2494	1	7	4.37	1.57	3704	1	7	4.61	1.48	3418	1	7	4.38	1.58
Item 2	2491	1	7	4.50	1.66	3699	1	7	4.61	1.53	3419	1	7	4.31	1.61
Item 3	2495	1	7	3.68	1.66	3708	1	7	4.11	1.66	3424	1	7	3.91	1.69
Self-handicapping															
Item 1	2488	1	7	3.75	1.59	3708	1	7	4.17	1.56	3429	1	7	4.28	1.62
Item 2	2492	1	7	3.59	1.53	3705	1	7	3.86	1.55	3416	1	7	3.90	1.70
Item 3	2485	1	7	3.75	1.61	3705	1	7	3.77	1.57	3422	1	7	3.71	1.62
Causal beliefs (luck)															
Item 1	2484	1	7	2.50	1.45	3705	1	7	2.41	1.50	3423	1	7	2.32	1.42
Item 2	2492	1	7	1.75	1.31	3703	1	7	2.03	1.48	3425	1	7	1.96	1.44
Item 3	2491	1	7	2.39	1.39	3702	1	7	2.80	1.53	3432	1	7	2.58	1.48