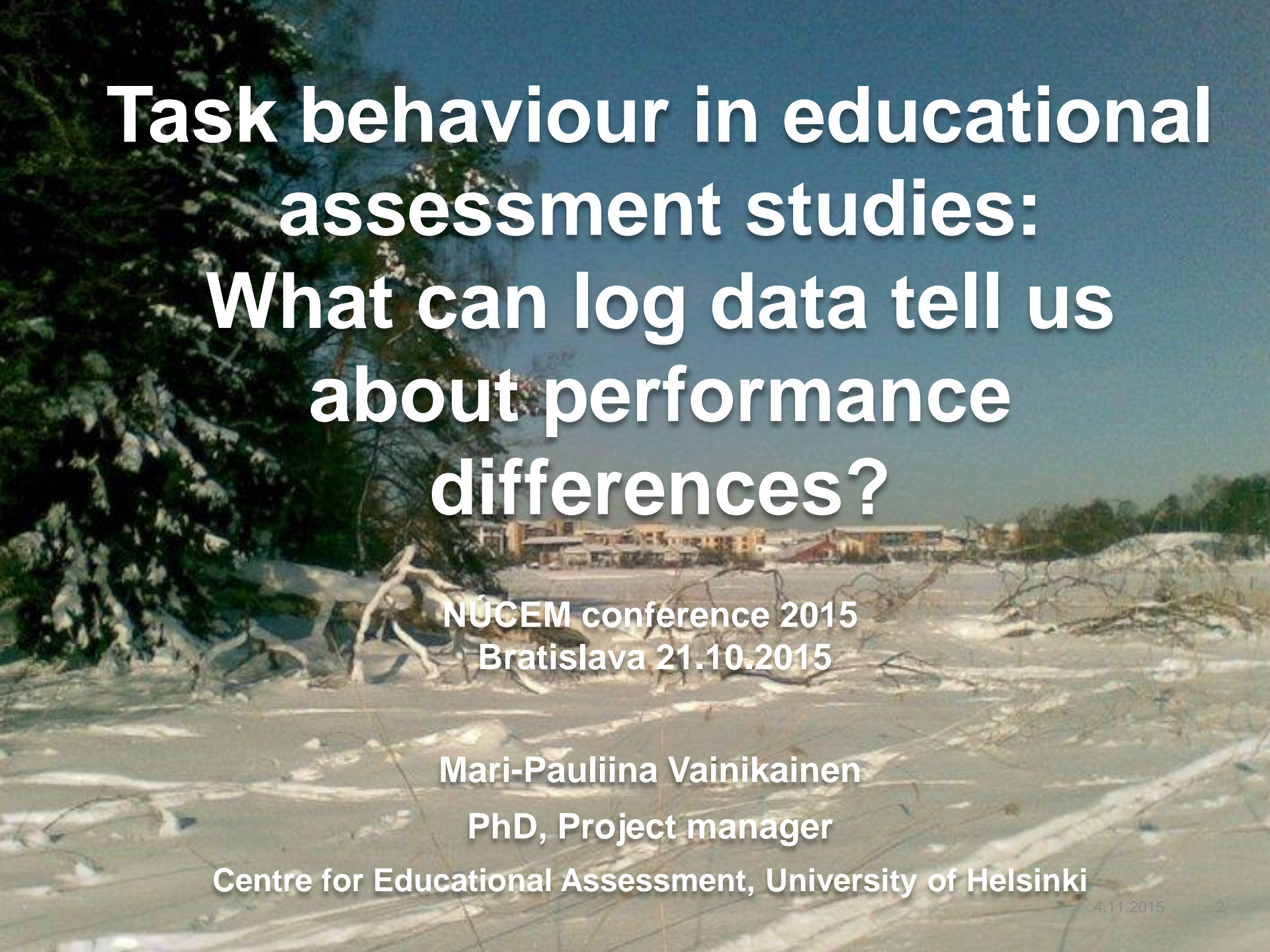


Task behaviour in educational assessment studies: What can log data tell us about performance differences?

**NÚCEM conference 2015
Bratislava 21.10.2015**

**Mari-Pauliina Vainikainen
PhD, Project manager**

Centre for Educational Assessment, University of Helsinki

A scenic view of a beach with driftwood and buildings in the background. The foreground is filled with light-colored sand and scattered pieces of driftwood. In the middle ground, there are several buildings, possibly a resort or a small town, situated along the coast. The background shows a clear blue sky and some trees on the left side.

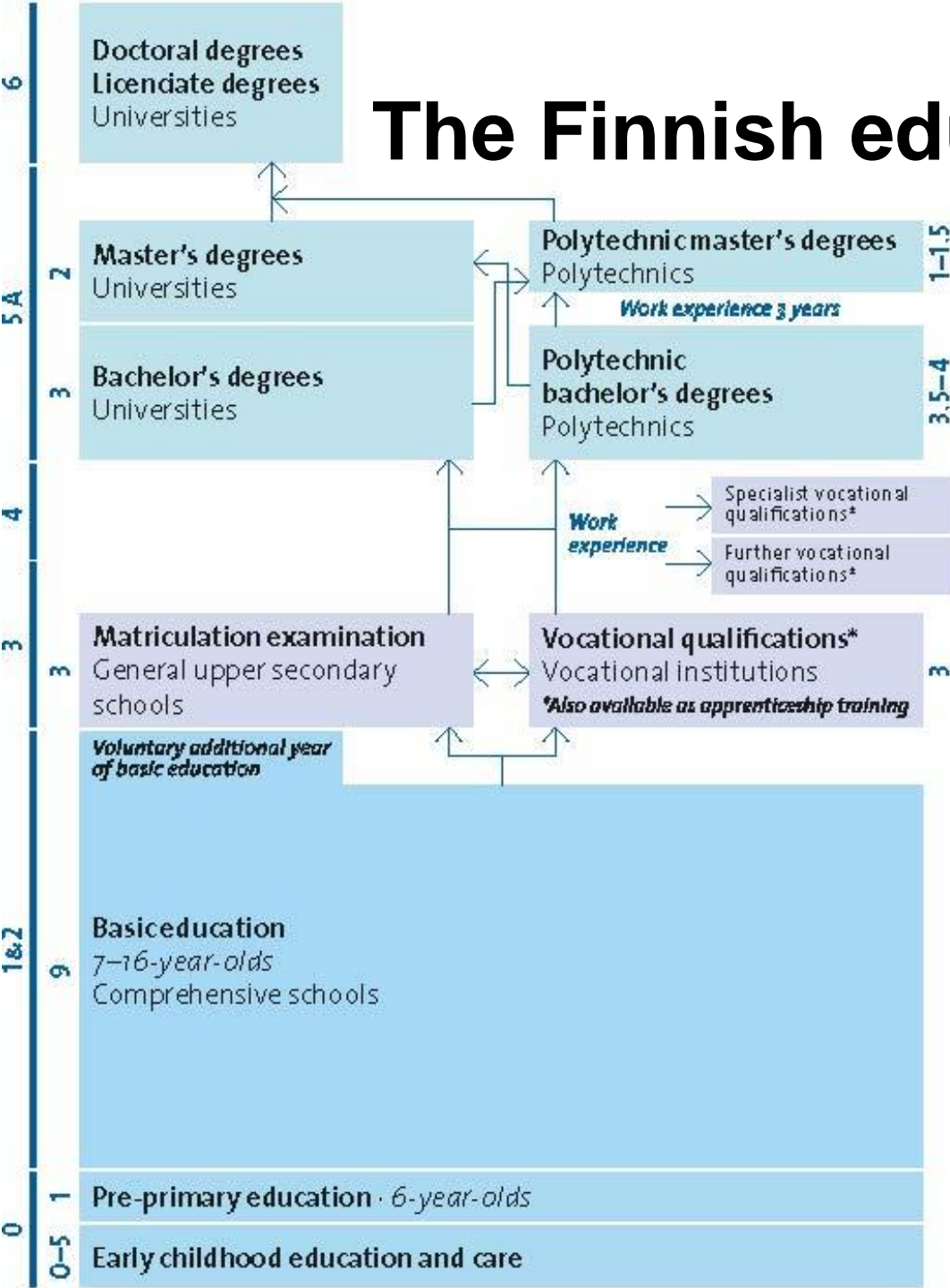
Task behaviour in educational assessment studies: What can log data tell us about performance differences?

**NUCEM conference 2015
Bratislava 21.10.2015**

**Mari-Pauliina Vainikainen
PhD, Project manager**

Centre for Educational Assessment, University of Helsinki

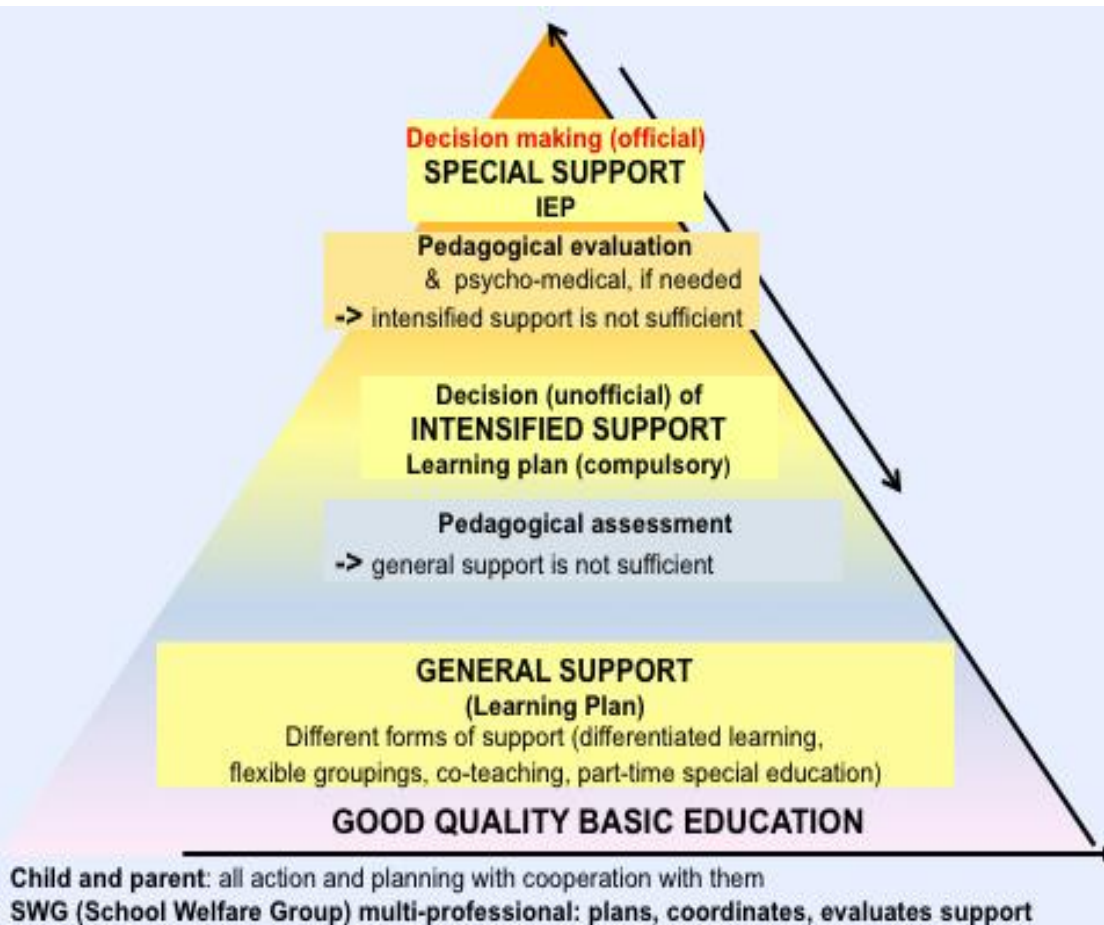
The Finnish educational system



- The Finnish basic education system is based on the idea of equity of education (Aho et al., 2006).
 - Also the results of large-scale international comparisons support this view (Willms, 2010).
- -> No tracking and very little school selection until the end of lower secondary education.



The three-tiered support model



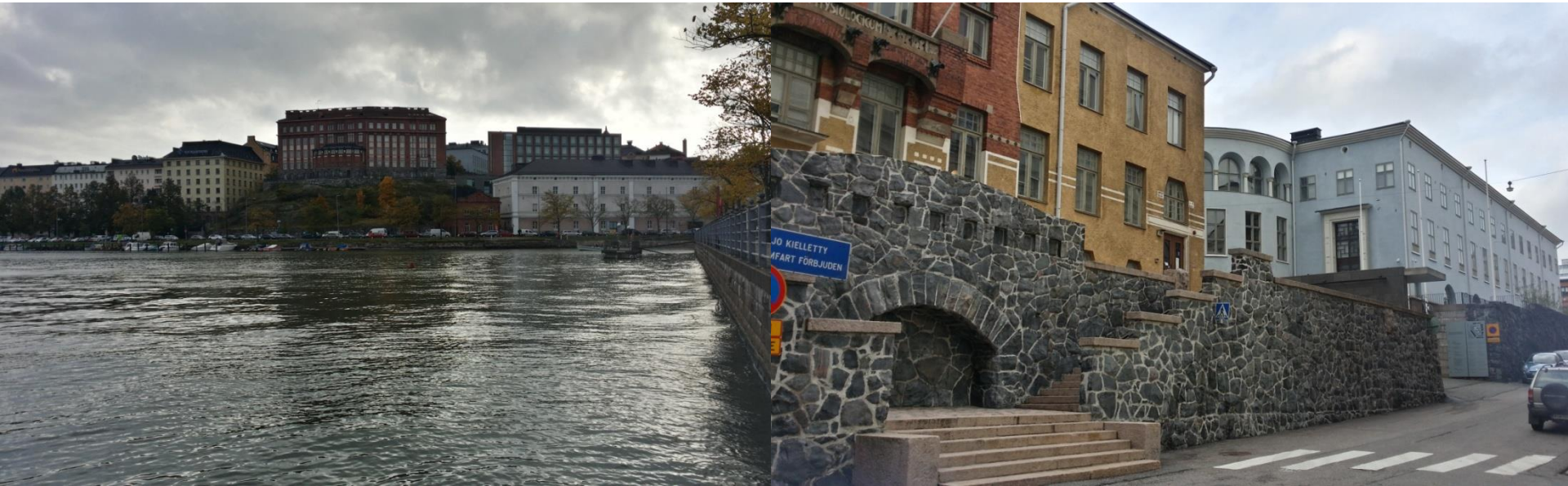
- One of the key components in securing the equity of education in Finland has been the support provided for the weakest learners (Graham & Jahnukainen, 2011; Sabel et al., 2011).
- Gender differences, however, have in Finland been larger than in many other countries (e.g. OECD, 2013), with girls performing better.



Centre for Educational Assessment



- The Centre for Educational Assessment at the University of Helsinki has since 1996 developed measures for assessing learning to learn (LTL) and to implement large-scale national and municipal assessment projects.





Learning to learn assessments



- Assessment of more general learning, thinking and problem-solving skills, and affective factors, which influence
 - school achievement
 - preparedness for life-long learning
- The targets of the assessment
 - cognitive competences (thinking, reasoning and problem solving skills in different contexts)
 - learning-related attitudes, beliefs and self-concepts



Centre for Educational Assessment



- Currently the centre works on several longitudinal computer-based large-scale assessment studies (N = 1000 – 14 000) on learning to learn and the development of learning outcomes and wellbeing.
 - The first computer-based assessments in 2001.
- The centre has recently developed a computerised adaptive test for student selection of vocational upper secondary schools.
- Involved in the national implementation of the computer-based PISA2015 and PISA2018 assessments in Finland.
- A new study for monitoring tablet computer use in one municipality that bought tablets for almost every student.



A short history of educational assessment in Finland



- The implementation of comprehensive education in the 1970s
 - Nine years of compulsory education for everyone free of charge
 - All the schools followed the same curriculum
 - School inspection and controlling of textbooks
- Loosening the control in the late 1980s and early 1990s
 - No more text book controls
 - The school inspection system was ceased
 - Curriculum reform in 1994: The national core curriculum as a basis for the local (municipal) curricula
- Introduction of the model for educational assessment.



The strategy for educational assessment in Finland 2012-2015

(Ministry of education and culture)



- The responsibility to assess the outcomes of education *locally* (= on a municipal level) is written in the Basic Education Act.
- Additional international and national *sample-based* low-stakes assessments
 - Their implementation has not been coordinated from one place so far.
- The only high-stakes test for the students (and only for those on the academic track of upper secondary education) is the matriculation examination, which is currently being digitalised.



International low-stakes assessments in Finland



- International comparative assessments
 - PISA, TIMMS, PIRLS, etc.
 - The results and the data could be used in a much more productive way than making ranking lists!
 - National reports for analysing the local trends and country-specific phenomena
 - National comparisons of different student subgroups (boys and girls, language groups, immigrants, etc.)
- Useful for policy and school development on a general level
 - Due to the restricted sampling very difficult to utilise as feedback for individual schools



National low-stakes assessments in Finland



- National assessments of the effectiveness of education
 - Sample-based assessments of students' skills in different school subjects
 - Every other year mother tongue and mathematics
 - Usually one more school subject in addition
 - Usually for 15-years-old students
 - A sample of about 10 % of the student population (150 schools)
 - The results are published only on a country- and region-level
 - > NO RANKING LISTS!
 - Schools are provided feedback for their own developmental work.

- "Thematic assessments", which are repeated more seldom
 - Learning to learn assessment as an example



Locally implemented educational assessments in Finland



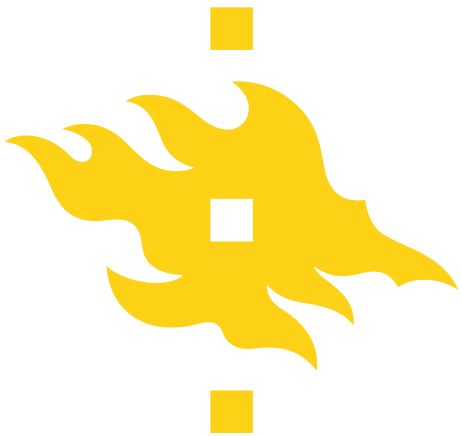
- Assessments organised by education providers or individual schools
 - Self-evaluation
 - Assessing the learning outcomes of all the schools in the municipality using the tests of the national assessments
 - Collaboration with universities, (e.g., learning to learn assessments)
- > Again, no rankings, the results are published only on a general level. The main aim with all this is to get information for school development. Individual teacher can use the results for their own professional development.
- > **ALL THE ASSESSMENTS ARE LOW STAKES FOR THE STUDENTS!**



Assessment of learning outcomes of individual students



- Teachers have the freedom to choose their own assessment methods used in grading and giving feedback
 - Summative assessment = measuring what a student has learned = Assessment OF learning
 - Criteria-based: The national core curriculum gives the description of good performance (Grade 8 on a scale of 4 to 10) in all school subjects
 - No national tests for this purpose – The consistency of grades?
 - With younger children even this type of assessment can be based on reaching the individual goals of each child
 - Formative assessment = Giving the student data-based feedback for enhancing learning = Assessment FOR learning
 - The feedback loop between the student and the teacher
 - The student needs to know exactly what is expected and how far the goal is from the current situation
 - The goals are always set individually
 - The role of the teacher is crucial!



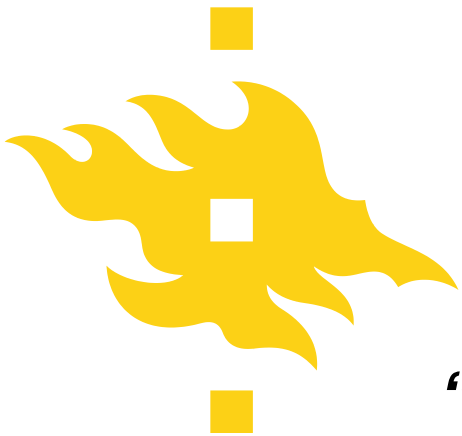
Why to study task behaviour?



Our worry...



- During the recent decade, the performance of Finnish 15-year-olds has been steadily declining in both international and national assessment studies
(see Hautamäki, Kupiainen, Marjanen, Vainikainen & Hotulainen, 2013, for a review)
- > One possible reason for the decline and also for the increasing differences between student subgroups is the simultaneously declining motivation and reduced effort of some students in low-stakes assessments.

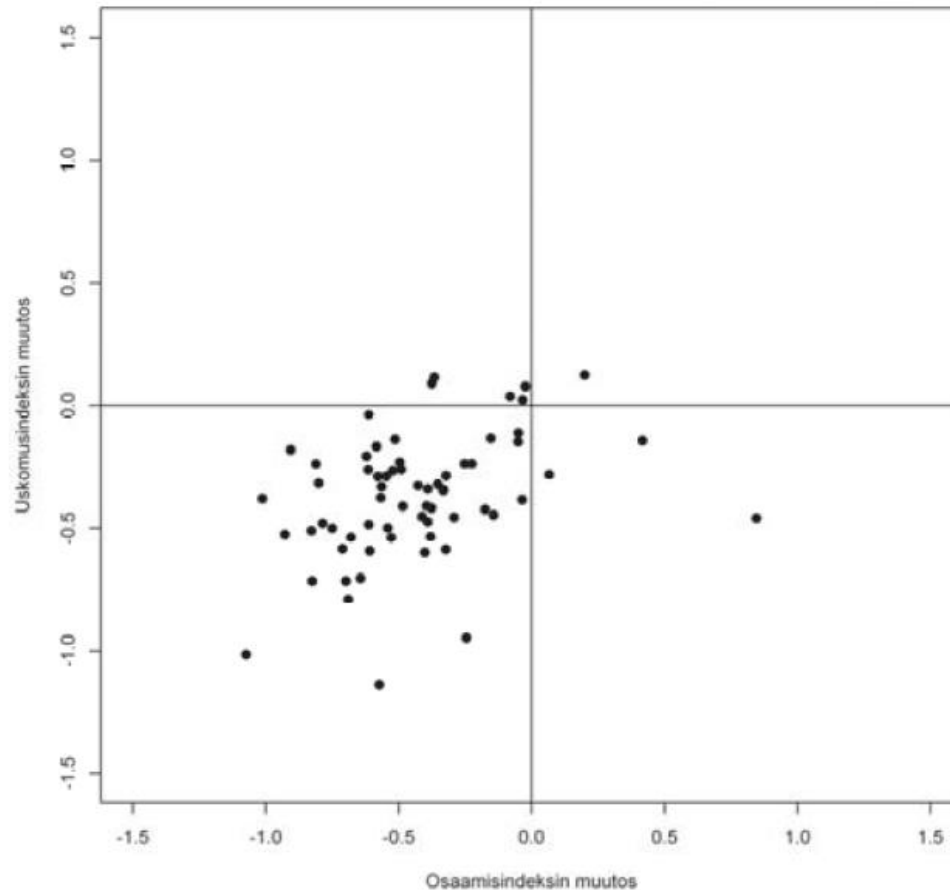


“Participating in a particular task requires the demonstration of the characteristics associated with the task, and whether this requirement is seen as an opportunity or a burden depends on the individual’s needs, explicit and implicit motives and personal values, and on the individual’s desire to demonstrate these characteristics both to herself and to others.”

Eccles, 2011



School-level development of cognitive LTL competences and motivational beliefs 2001-2012





Why to study task behaviour in CBA?



- It is an internationally acknowledged problem that the results of educational assessments may be influenced by reduced effort if the assessments do not have any personal consequences for the students (low-stakes assessment).

(Eklöf, Japelj Pavešič, & Grønmo, 2014; Setzer, Wise, van den Heuvel, & Ling, 2013; Wise, 2006)

-> We need measures for controlling the effect of effort and to analyse, to what extent the observed group differences in performance are deriving from non-cognitive factors.



Measuring effort in PBA and CBA



- In paper-based assessments, effort has traditionally been measured by including self-report questionnaires in the assessment batteries (e.g., OECD, 2013).
 - This has been shown to be relatively unreliable (Wise & Kong, 2005).
- Log data analyses of time investment in CBA has been demonstrated to be a much more accurate measure of effort (Wise & Kong, 2005).



Gender differences in performance, motivational beliefs and effort



- There should be no gender differences in general cognitive competences even though there are some differences in underlying specific abilities (Halpern, 2000).
 - Girls perform better in verbal tasks
 - Boys perform better in visuo-spatial and mathematical tasks
- Nevertheless, girls often outperform boys both in school subjects (e.g. Kimball, 1989) and in external assessments (e.g. OECD, 2013).
- Girls have been found to have motivational beliefs and goals, which support learning better (Kenney-Benson et al., 2006).
- Girls have also been found to be more persistent when facing an uninteresting task (Ainley et al., 2002).



Support needs, performance, motivational beliefs and effort



- Obviously, support needs are often related to lower cognitive competences.
 - An indicator of effectiveness of support could therefore be that the differences would not increase over time.
- Pupils with support needs are often low-achievers who also have low motivation (= lower effort?) (Thuneberg, 2007).
- More boys than girls need support for their studies
(Halpern, 2000; Thuneberg, 2007)
- Due to differences in how support needs are identified and categorised, there is little evidence from large-scale assessment studies in how pupils in need of support are performing compared to the others.



Tradition of time on task



- Carroll suggested already in 1963 that learning is determined by the ratio of the time needed and the time spent on learning.
 - The *time needed* depends on students' initial competences, their ability to understand instruction, and the quality of instruction.
 - The *time spent* depends both on the time allocated for the assessment and the time an individual pupil is willing to spend on them.
- When time investment in an assessment setting is studied, it is thus important to take into account prior competences in order to relate the invested time to the **actual needs of the student.**



Time on task research nowadays



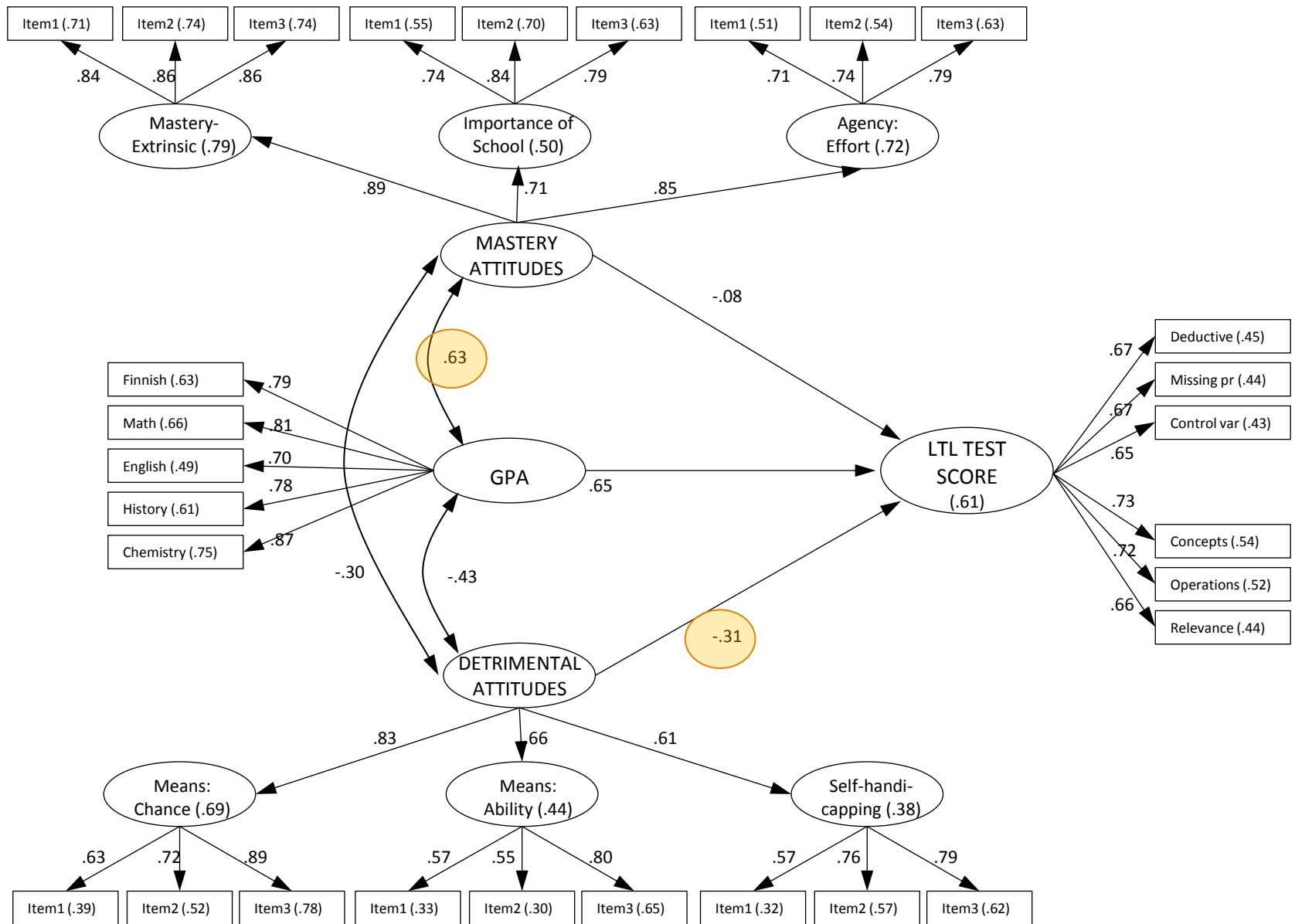
- Often item-based log data analysis
 - In high- vs. low-stakes test situation comparisons, it has been noted that in high-stakes testing, *rapid guessing* (see Schnipke & Scrams, 1997) occurs typically towards the end of the assessment session when the time is running out, whereas in low-stakes testing insufficient time investment has been spread much more evenly throughout the assessment session (Wise & Kong, 2005).
- We have used task-based log data of 8-10 items displayed together to study the effects of motivational beliefs on time investment.



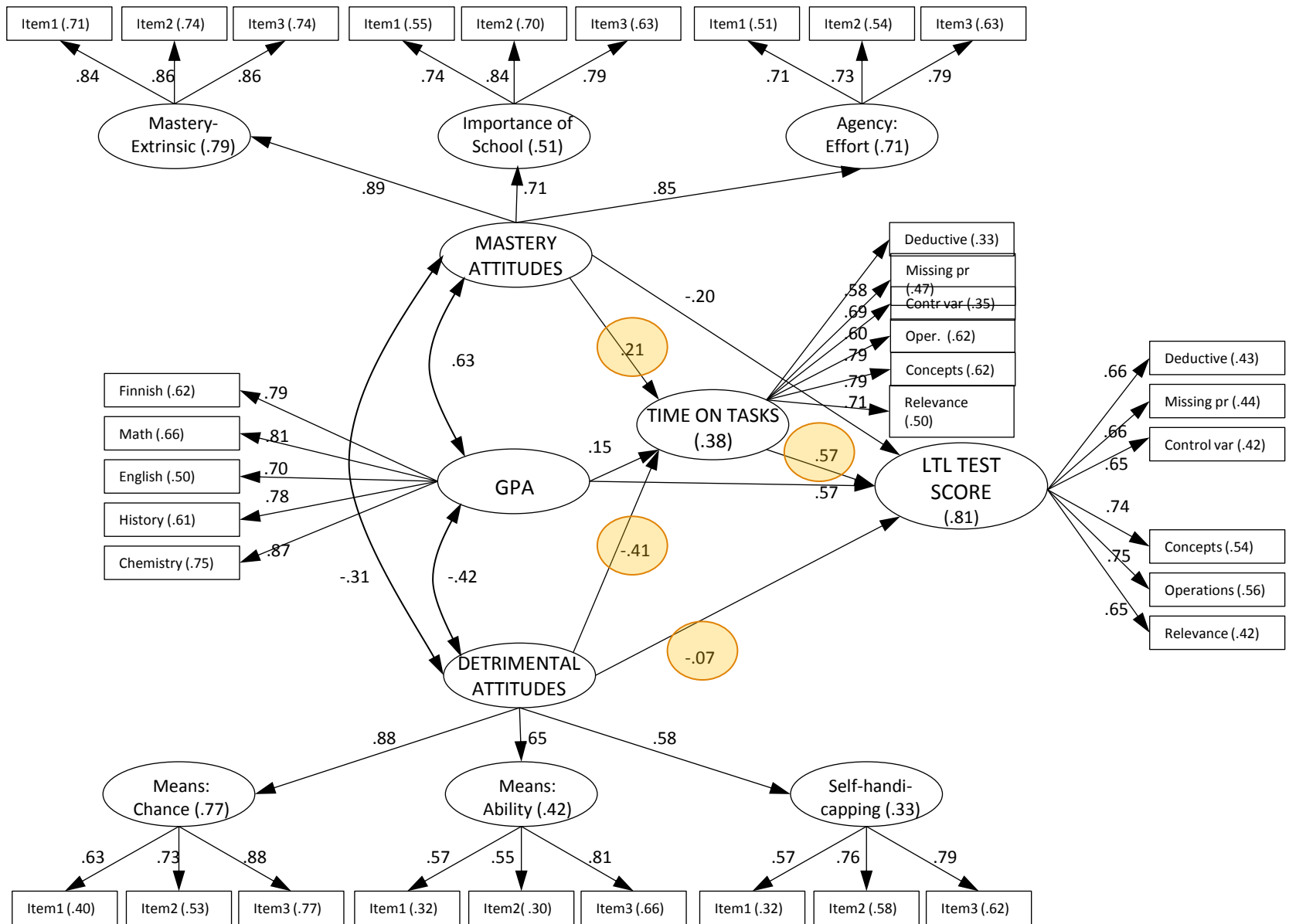
Time on task research nowadays



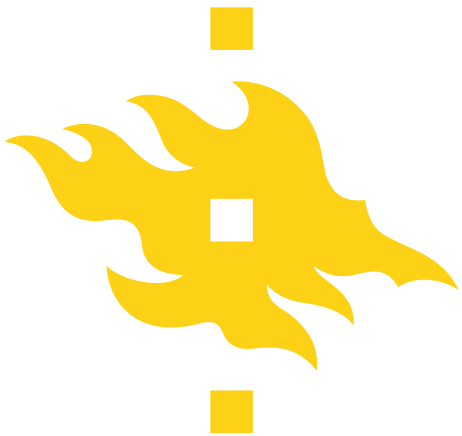
- Task type matters:
 - In routine tasks that require a high level of automatization, shorter time on task has been associated with better performance (Goldhammer, Naumann Stelter, Tóth, Tölke, & Klieme, 2014).
 - In tasks that require reasoning and problem-solving even better performers benefit from increased time investment (Goldhammer et al., 2014; Kupiainen et al., 2014).
- Motivation and interest predict time investment (Ainley et al., 2002a; Kupiainen et al., 2014; Scherer, Greiff & Hautamäki, 2015) and more generally task behaviour (Timmers et al., 2013).



Source: Kupiainen, S., Vainikainen, M.-P., Marjanen, J., & Hautamäki, J. (2014). The role of time on task in low stakes assessment of cross-curricular skills. *Journal of Educational Psychology*, 106 (4).



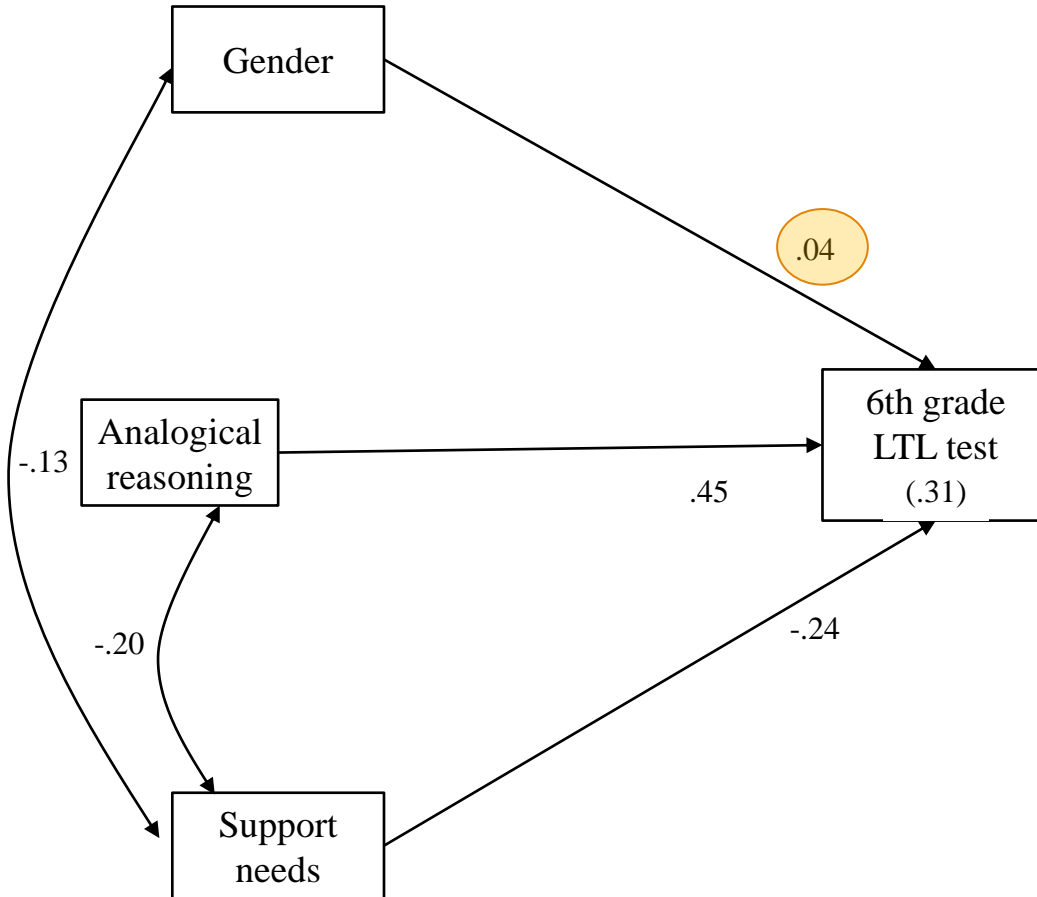
Source: Kupiainen, S., Vainikainen, M.-P., Marjanen, J., & Hautamäki, J. (2014). The role of time on task in low stakes assessment of cross-curricular skills. *Journal of Educational Psychology*, 106 (4).



Explaining group differences in performance

Source: Vainikainen, M.-P. (2014). *Finnish primary school pupils' performance in learning to learn assessments: A longitudinal perspective on educational equity*. University of Helsinki, Department of Teacher Education Research Reports 360. Helsinki: Unigrafia.

Gender and support needs in predicting the 6th grade test performance



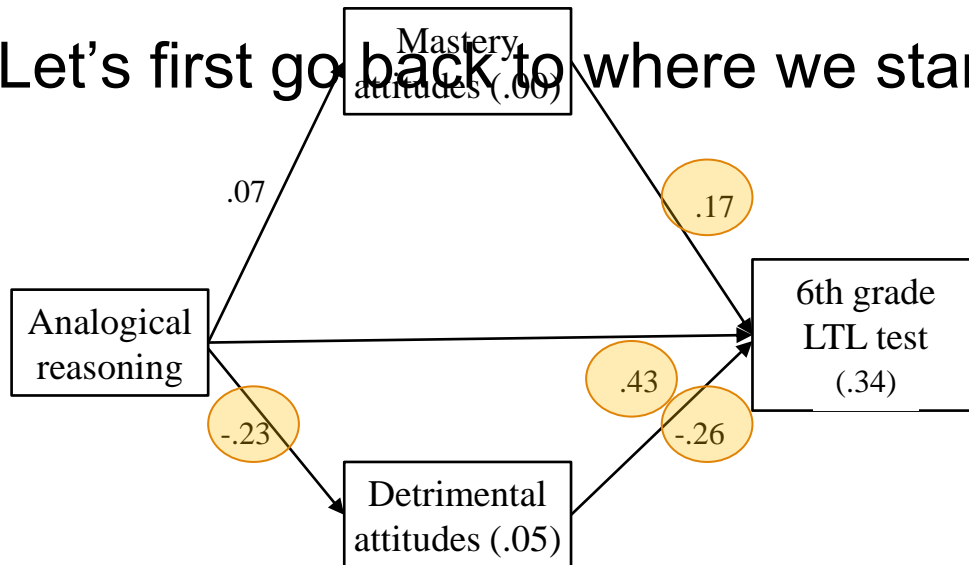
- There was no gender difference in 3rd grade analogical reasoning
- Girls performed slightly better in the 6th grade LTL test (M=41.36 vs. M=38.58)
- Support needs were related to lower performance in the analogical reasoning test and the LTL test
- Boys had more support needs



■ CFI=.997, TLI=.975, RMSEA=.030, $\chi^2=2.396$, df=1, p=.122

Motivational beliefs in predicting test performance

Let's first go back to where we started.

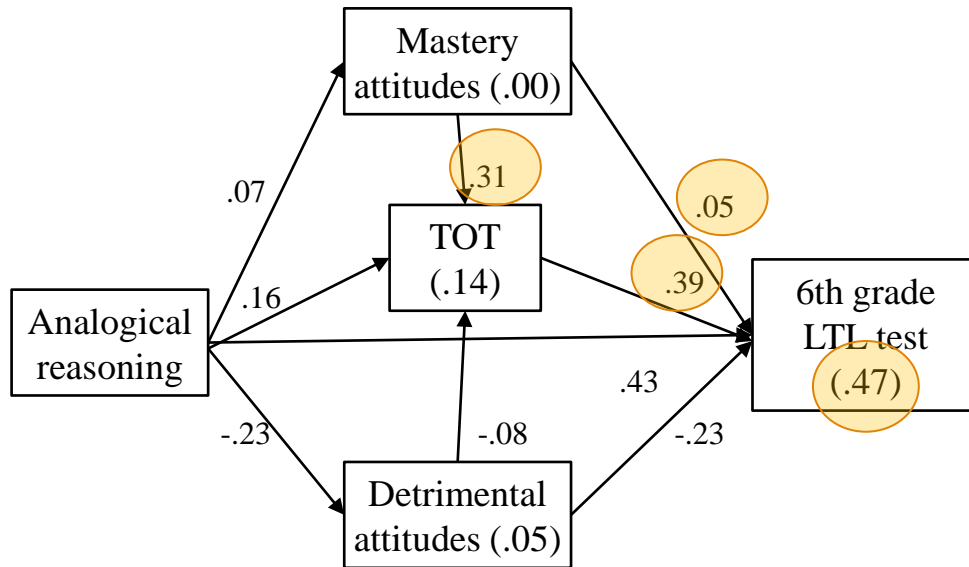


- Mastery attitudes predicted performance positively and Detrimental attitudes negatively
- Lower analogical reasoning skills predicted higher levels of detrimental attitudes
- Detrimental attitudes partially mediated the effect of analogical reasoning



■ CFI=.995, TLI=.955, RMSEA=.044, $\chi^2=3.931$, $df=1$, $p=.047$

Time on task and motivational beliefs in predicting test performance



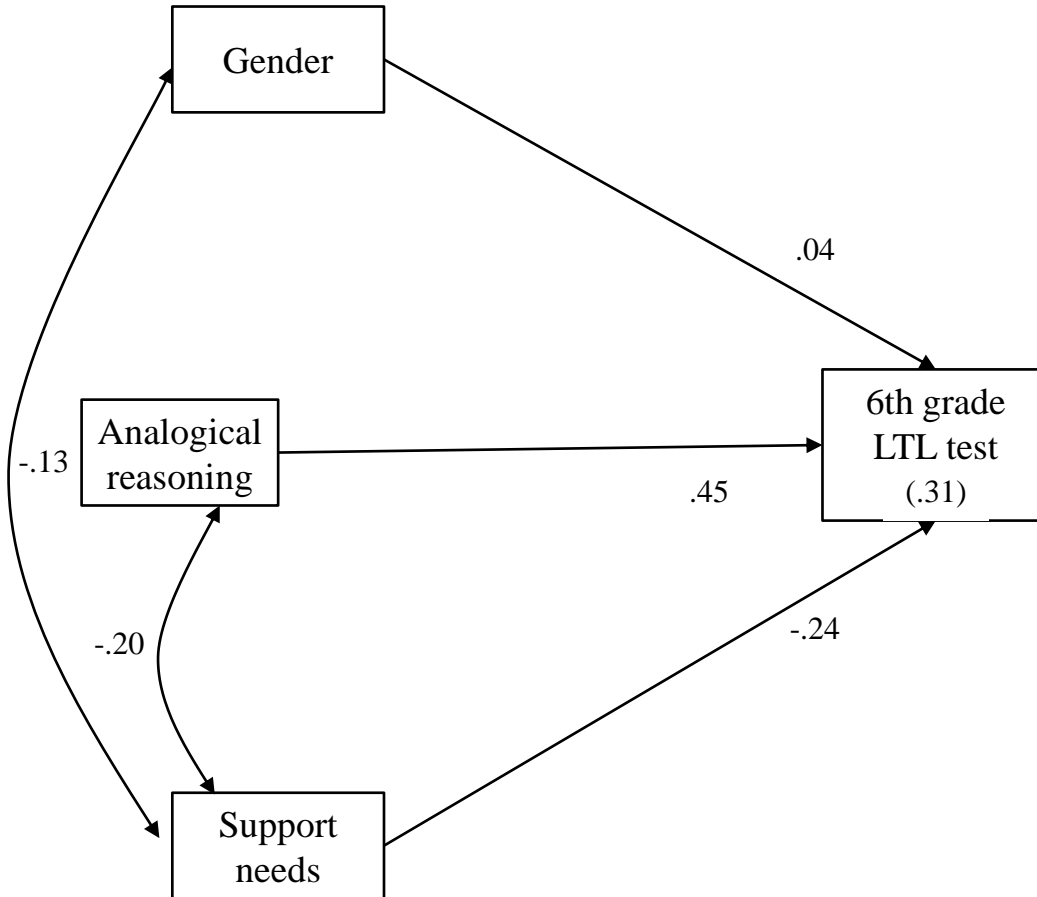
- Time on task mediated the effect of Mastery attitudes almost completely
- Time on task was a relatively strong predictor of performance in the 6th grade LTL test



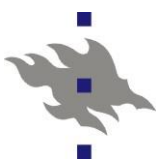
- CFI=.997, TLI=.962, RMSEA=.044, $\chi^2=4.047$, $df=1$, $p=.044$



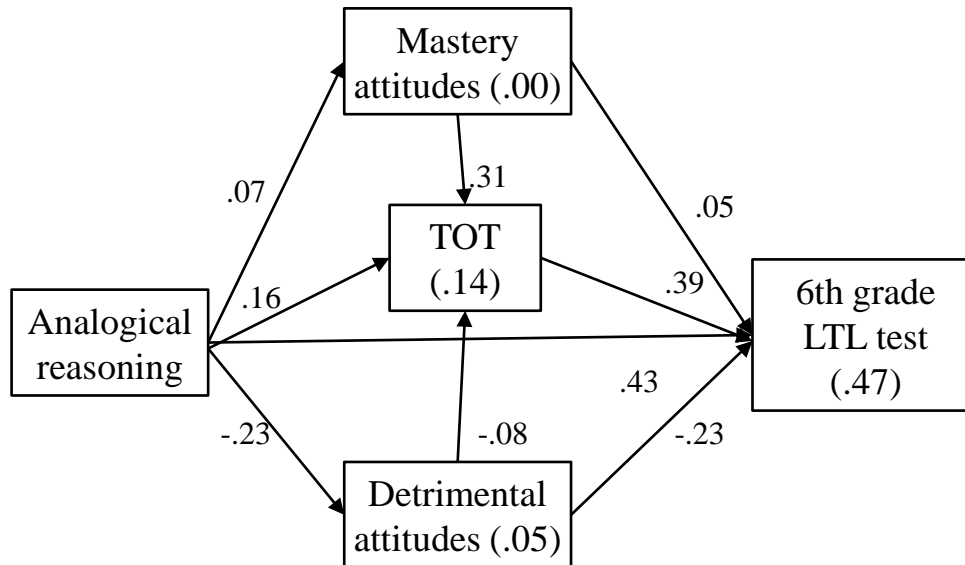
Finally, what happens if we combine this model...



- CFI=.997, TLI=.975, RMSEA=.030, $\chi^2=2.396$, $df=1$, $p=.122$



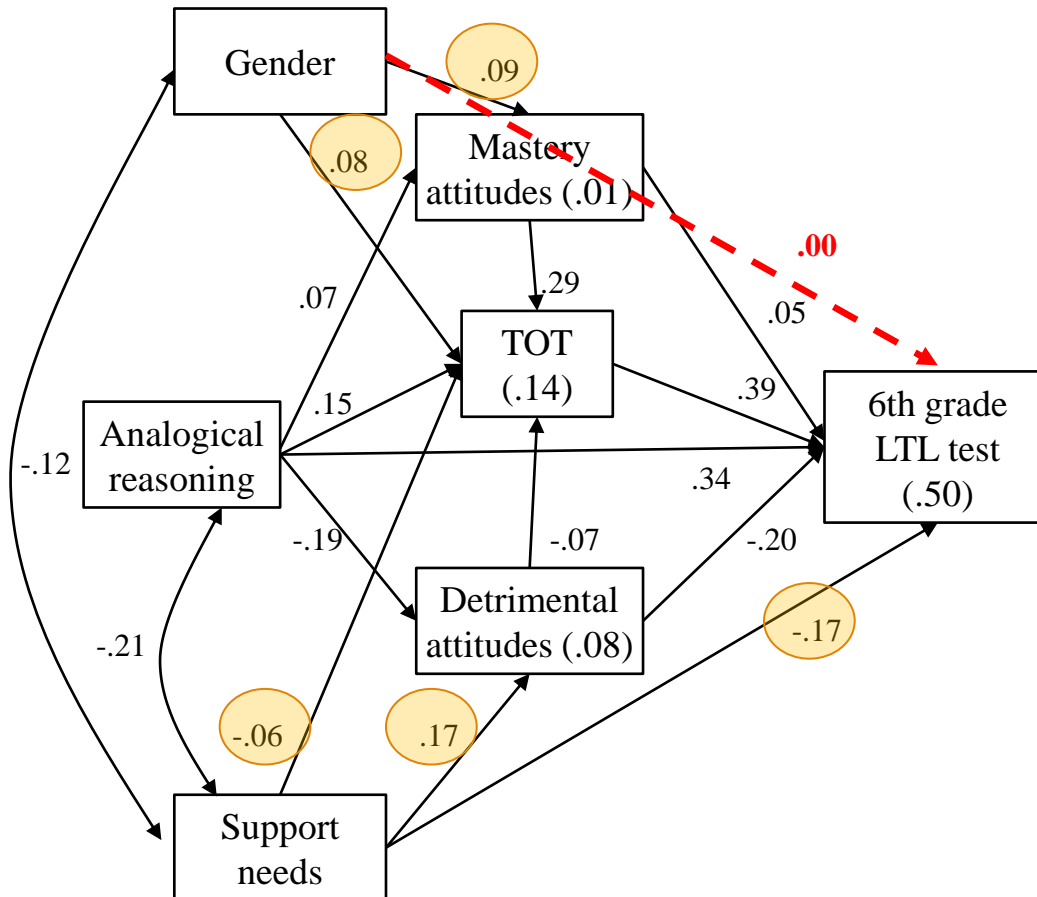
... with this model?



- CFI=.997, TLI=.962, RMSEA=.044, $\chi^2=4.047$, $df=1$, $p=.044$



The final model for predicting the 6th grade LTL test score



- Gender no longer predicted performance directly.
- Instead, girls had more Mastery attitudes. They also spent more time on the assessment tasks.
- Support needs predicted reduced Time on task and a higher level of Detrimental attitudes



Yet, support needs still predicted the test score directly ($\beta = -.17$ vs. $\beta = -.24$)

- CFI=.996, TLI=.978, RMSEA=.027, $\chi^2=10.753$, $df=5$, $p=.057$



CONCLUSIONS

- The small but statistically significant gender difference could be completely explained by the other variables in the model
 - Girls had less support needs (cf. Thuneberg, 2007).
 - Girls had more Mastery attitudes (cf. Kenney-Benson et al., 2006).
 - Girls spent more time on tasks (cf. Ainley et al., 2002).
- The somewhat larger effect of support needs on performance could only partially be explained by the other variables in the model
 - Support needs predicted a higher level of Detrimental attitudes (cf. Thuneberg, 2007).
 - Support needs predicted a **reduced** time use (vs. Carroll, 1963).



CONCLUSIONS

- The "boy problem" in Finland seems to be at least in the 6th grade mainly a matter of motivational attitudes, effort and persistence
(cf. Ainley et al., 2002).
 - The studies should be replicated with other learning to learn datasets and using other measures (e.g. curriculum-based assessments, PISA)
 - Implications for interventions?
- Log data analysis of task-based time investment is a relatively simple way of controlling test-taking effort in a low stakes assessment situation.
 - The future course of research is to pay more attention to other aspects of task behaviour like problem-solving strategies.



KIITOS

mari-pauliina.vainikainen@helsinki.fi



The measured concepts in the learning to learn assessments



- COGNITIVE COMPETENCES
 - Reasoning skills
 - Reading comprehension
 - Mathematical thinking
 - Problem-solving

- ATTITUDES AND BELIEFS
 - Motivational beliefs
 - Academic self-concepts
 - General self-concept
 - Commitment to school work
 - Feelings related to the environment (School, class, teachers)
 - The support from significant others (Parents, friends, peers)



The structure of the thinking skills we measure



- Our recent research (Vainikainen, Hautamäki, Hotulainen & Kupiainen, 2015) shows that the thinking skills we measure are organised hierarchically
 - There are both general and domain-specific (yet independent of the curricular contents) mechanisms that are required for thinking and problem solving
 - The cognitive tasks we currently use address the verbal and the quantitative specialised structural systems (using the terminology of Demetriou et al., 2011) in addition to the general mechanisms. According to the theory, there are specialised systems for spatial, categorical, causal, and social thinking too.



The structure of the thinking skills we measure



- As originally claimed in the theoretical Finnish learning to learn framework, the study proved empirically that the general thinking skills factor was dominated by Piagetian formal thinking.
- -> The importance of the developmental approach also when it comes to learning to learn and thinking skills: The progress to the next level of thinking goes stepwise, not always in a linear way.