The effects of opportunity to learn on students' achievement: a meta-analysis

Master thesis

Marloes Lamain
Educational Science and Technology

UNIVERSITY OF TWENTE.

The effects of opportunity to learn on students' achievement: a meta-analysis

AUTHOR
Marloes Lamain
Student number: s1494910
Email: marloeslamain@gmail.com

SUPERVISORS
First: Dr. J.W. Luyten (Hans)
Second: M.R.M. Meelissen

DATE
May, 2018

Acknowledgement

It has given me pain and trouble, but now I am happy and proud that I can present my master thesis. It’s the final assignment I had to finish in order to complete the master Educational Science and Technology at the University of Twente.

In August 2015 I was asked to participate in a research team as a junior-researcher under the direction of Jaap Scheerens. I found this a great opportunity from which I have learned a lot. The research was financed by the Netherlands Initiative for Educational Research in the Hague. The study made a step towards updating the state of art concerning opportunity to learn, by giving a conceptualization, by taking a closer look at meta-analyses and by performing a new meta-analysis. My main task in this team was carrying out the literature search.

After we have finished the research ‘Opportunity to Learn, Curriculum Alignment and Test Preparation’ I used the literature search and the obtained information for my master thesis. This was easier said than done. I was allowed to use the same data, but not to conduct the same research. It was a puzzle succeeding a thesis with the same data, but with a different core. Now it is finished, I think I can say that I added useful information by further deepening into the data. I would like to thank the team members from the original study and my supervisors:

- Jaap Scheerens, because he gave me confidence during my first ‘real’ research by always complimenting my work.
- Peter Noort, because I could ask him all my questions, none were too crazy for him.
- Hans Luyten, team member and my supervisor, because he stimulated me to continue working on my thesis at times when I doubted to give up.
- Martina Meelissen, my second supervisor, for giving me a final portion of well thought-out feedback.

Besides them, I want to thank my parents for letting me live in their house again, so I did not have to worry about financial aspects like a proper job and paying rent. They gave me the opportunity to obtain two bachelor’s degrees and with this thesis, my master’s degree.

Marloes Lamain
Boekelo, March 2018

Abstract

Aligning the intended, implemented and attained curriculum can improve student achievement. This form of alignment is often seen as opportunity to learn (OTL). It is all about the opportunity the students get to learn the content, in accordance to the national standards, that will be tested. The relation between OTL and student achievement has been studied extensively, but it was not clear to with extent and under which conditions these are related. For this reason a meta-analyses about OTL was conducted by Scheerens et al. (2017). This thesis builds on this meta-analysis, by seeking an explanation for the modest proportion of statistically significant positive OTL effects that Scheerens et al. (2017) found. First, the definition of OTL was narrowed to one aspect of OTL: content coverage. However, this only caused a minimal improvement in the proportion of significantly positive effects. Second, the influence of study characteristics was examined by multiple chi-square tests. Again, without a clear result. Only the study characteristic 'subject' showed a significant association, but in the opposite direction than expected: effects in studies with mathematics as variable are less likely to be significantly positive than effects in studies with other subjects as variable. As a result of this study it can be concluded that both, a narrower definition of OTL, as well as, the influence of study characteristics, gave no explicit explanation for the modest proportion of significantly positive OTL effects in Scheerens et al. (2017).

Keywords: Opportunity to learn, alignment, content coverage, student achievement

Table of contents

List of graphs and tables

Tables

Crosstabulations

Graphs

Introduction

Problem statement

The 28th article of the convention of the rights of the child declares that all children have the right to education. States parties shall make education obligatory and free to all (The United Nations, 1989, art. 28). A government should send every child to school for proper education. However, even with financial obstacles disregarded, that is easier said than done. Effective education does not arise spontaneously. Two well-known authors in the field of education wrote their own extensive meta-analysis about the effectiveness of education. Robert J. Marzano's book *What works in schools* (2003) describes his findings about factors affecting student achievement based on a review of education research over a 35-year period. Ostensibly things work in education, but according to Hattie (2009) a mind swap must be made from how to make things going to how to make things work best. John Hattie wrote his book *Visible Learning* in 2009 in which he ranked 138 aspects influencing learning outcomes. Two years later he updated his list to 150 influences and in 2015 even to 195. The extensiveness of his list shows the complexity of education. Despite the list based on almost 1200 meta-analyses, there is still a lot of discussion about how to achieve effective education. This study will focus on one of the expected effective aspects of education, namely 'Opportunity to Learn', abbreviated as OTL.

OTL is closely related to alignment between national standards (intended curriculum), the implemented curriculum and students' learning outcomes (attained curriculum). Alignment is generally accepted as an important condition for effective education as long as the target is adequate. Focus must be on the right content; content about which the last word not has been spoken yet (Porter, Smithson, Blank, & Zeidner, 2007). Better alignment between national standards and the implemented curriculum will lead to better learning outcomes. In literature this form of alignment is often called OTL. It is all about the opportunities the students get to learn the content in accordance with the national standards, that will be tested (Squires, 2012). Earlier research from inter alia the above described Marzano and Hattie, show positive effect sizes for OTL when it comes to educational effectiveness (Scheerens, 2015). Scheerens (2015) noted that the number of recent meta-analyses specifically about OTL is limited in his review of the research literature on educational effectiveness.

Original study: Scheerens, Lamain, Luyten, & Noort (2017)

Scheerens (2015) has argued that additional research is needed. His study wants to contribute by clarifying ''the complexity of alignment between curricular elements'', by providing ''suggestions for legitimate test preparation'' and by giving ''suggestions for placing OTL and instructional alignment on the agenda of task related teacher cooperation'' (Scheerens et al. 2017, pp. 2-3). One of the research questions the study addressed was: ''What is the average effect size of OTL (association of OTL with student achievement outcomes), as evident from available meta-analyses, review studies, secondary analyses of international data-sets and (recent) primary research studies?'' (Scheerens et al., 2017, pp. 3). To answer this question, several literature searches were done to find meta-analyses and review studies about OTL. Eventually Scheerens et al. (2017) found a proportion of statistically significant positive effects of 43.8%. A modest effect size, because it can be expected that students achieve better results as the tested content has received more attention.

Follow-up study

This study will try to find out whether the outcomes of the meta-analysis of Scheerens et al. (2017) could be influenced by the way his study has defined OTL and the study characteristics of the studies under review. Researchers interpret and define OTL differently, some include only content variables or time related aspects, while others include also school factors, teacher factors and personal factors. The definition of OTL of the studies in the meta-analysis included ways of teaching, types of curricula, types of textbooks and supplementary services. However, content coverage was the most common OTL variable, therefore, this study will use a more precise and smaller definition of OTL by including only OTL effects concerning content coverage.

Secondly, this study will explore the possible effects of study characteristics on the outcomes of the meta-analysis. A total of six study characteristics are examined. These are year of publication, geographical area, subject, type of education, type of respondents for data collection and number of respondents. Reasons for further investigation into these variables are described below.

1. Year of publication

In recent decades schools experienced increasing pressure from the authorities (Perryman et al., 2011). Schools have gained more and more freedom to spend government money the way they want, but at the same time they must account for their efforts and results (Bronneman-Helmers, 1999 & Zeichner, 2010). A lot of attention is nowadays paid to the effectiveness of education of which OTL is one aspect. This increased attention may have led to more goal-oriented teaching on a worldwide level. This may have let to smaller differences and therefore a smaller OTL effect through the years.

2. Geographical area

More than other continents American education is associated with grade inflation. Grade point averages in America have been increasing for decades. Stroebe (2016) states that this is partly the result of grading leniency owing to the great importance that is attached to student evaluations of teaching. Teachers want a high evaluation score and this encourages them to change their way of teaching; they teach their students the way they think their students want to be taught, including little work, entertaining classes and high grades (Crumbley, Henry, & Kratchman, 2001). This national trend of grade inflation can result in lower OTL effects compared to research in continents where grade inflation is less apparent. Therefore, in continents other than America more reliable grades are expected, resulting in a stronger OTL effect.

3. Subject

In educational research achievement in mathematics is often used as the dependent variable. The meta-analysis of Scheerens et al. (2017) included 38 studies of which 26 used only mathematics as the dependent variable. A possible reason for this is clarity in testing, an answer on a test item is often right or wrong, because of common use of multiple choice questions and closed open questions. The latter category means that the question is formulated in such a way that just one answer is correct (Verhage & de Lange, 1996). Assessing other subjects, like science, history or language can be more complicated, because of more frequent use of open questions; the answer is not always right or wrong but it can be partly correct. Besides this, mathematical achievement is assumed to be little influenced by other factors than school (and homework), while for example subjects like reading literacy can also develop strongly at home (Reeves, 2012). These are important differences between mathematics and other subjects resulting in expected stronger OTL effects for mathematics.

4. Type of education

Differences between primary and secondary education are hard to describe, because this can differ per country. Because the vast majority of the studies are conducted in North-America and most other countries have similar systems, primary and secondary education in North-America is further compared. Primary schools mostly have one or two permanent teachers per group of pupils, who teach their lessons in the same classroom. Whereas secondary schools have specialist teachers, who teach only one or two subjects to different groups of students. Due to this, a primary teacher has more opportunities for subject integration, where aspects from different fields of education are offered in an integrated way (Carr, 2007). A primary school teacher sees his pupils a lot and therefore knows well what content his students are exposed to. On the other hand, a specialist teacher in secondary education might be better in translating learning goals into proper instruction. Because there are arguments favouring primary education, as well as arguments that are favouring secondary education it is unclear which type of education a stronger OTL effect can be expected.

5. Type of respondents for data collection

Teachers and students are both important actors in education, but they do not always think the same about educational issues. Responses of students about instruction are often biased (Heafner & Fitchett, 2015). There is a bigger chance of socially desirable responses of children compared to responses of adults. This is due to the fact that their self-concept and attitudes are not fully developed yet (Butori & Parguel, 2012). Besides, students can easily forget to what content they have been exposed to. Therefore, having students as respondents can give different results compared to studies where teachers are the respondents. The degree of difference depends on matters like anonymity, consequences of study results and their relationship with the researcher. Striking is, for example, the comparison between the TIMSS 2011, based on teacher responses, and the PISA 2012, based on student responses, which revealed the OTL effect being much stronger in the PISA study compared to the TIMSS study (Scheerens et al, 2017). By contrast, a study of Herman, Klein and Abedi (2000), has found high correlations between student- and teacher-reported measures of OTL. To examine if there were reasons for students and teachers to participate differently in the studies used for this research, this study characteristic will be examined with regard to the proportion of statistically significant positive OTL effects found in Scheerens et al. (2017). A stronger OTL effect is expected for the studies in which the data comes from students. The content they indicate as covered is probably the same as the content in test items they answer correctly, because that is the content they remember.

6. Number of respondents

There is a higher risk of a publication bias in studies with a small sample size (Schwarzer, Carpenter, & Rücker, 2015). In case of a publication bias, small studies with positive results have a greater chance of being published, in contrast to studies with a negative or a vague result. However, for scientific research both positive and negative results are important. By investigating this study characteristic, it becomes clear if there were studies with a publication bias, when the OTL effect is stronger in small studies.

Research questions

The study of Scheerens et al. (2017) gives a clear overview of the number of articles satisfying a certain characteristic. This thesis goes one step further than the original study by looking at the influence of a certain characteristic on the proportion of statistically significant positive effects. Is there a difference between studies concerning mathematics versus studies concerning other subjects? Does it matter if the study is conducted in a primary school instead of a secondary school? What happens when the number of participants increases? These types of questions are being addressed in this study. In addition, this study will also examine if a narrower definition of OTL will contribute to a higher proportion of statistically significant positive effects. This leads to the following research questions:

Which study characteristics correlate with the proportion of statistically significant positive effects of OTL in Scheerens et al. (2017)?

A. *To what extent does the proportion statistically significant positive effects change when including only OTL effects concerning content coverage?*

B. *What is the influence of study characteristics (year of publication, geographical area, subject, type of education, type of respondents and number of respondents) on the proportion of statistically significant positive effects of the OTL variable content coverage?*

Overview of this thesis

The thesis is structured as follows: In chapter 2 the theoretical framework is described, in which OTL is explained by looking at how the concept has developed through the years, by discussing earlier research and by examining OTL from the perspective of different research traditions. The methodology for this study is explained in chapter 3. The next chapter, the fourth, shows the results, followed by the conclusion and discussion in chapter five.

Theoretical framework

Opportunity to learn

Opportunity to learn (OTL) is not a new concept, even though nowadays much more is written about OTL than a few decades ago. In studies published before 1980, the definition of OTL is quite narrow with a focus on the similarities between the content that has been taught and the test content. In the following decades, broader definitions were proposed including not only the content that is taught, but the way it is taught and by whom it is taught as well (Elliott, 1998). Even learner variables, like if a student has a computer at home, are sometimes taken into account. This diversity can cause confusion about the definition of OTL.

Stevens (1993) developed a framework for OTL which included four OTL components, retrieved from earlier research. The four elements in this framework are content coverage, content exposure, content emphasis and quality of instructional delivery. These terms have the following meaning:

- Content coverage: Access to the core curriculum and alignment between the curriculum that has been taught and the test content.
- Content exposure: How much time the students have to learn the concepts and skills (in-depth teaching).
- Content emphasis: The extent to which teachers select and emphasize certain topics.
- Quality of instructional delivery: The teaching practice including teachers' cognitive command, the use of varied teaching strategies and practices and coherent lessons. (Stevens, 1997; Stevens, Wiltz, & Mona, 1998; Herman & Abedi, 2004; Boscardin, Aguirre-Muños, Stoker, Kim, & Lee, 2005)

Every element in turn can consist of several variables. The clear structure in the framework has been a guidance for many researchers (Kurz, Elliott, Kettler, & Yel, 2014).

From all the OTL definitions used by researchers, content coverage is most discussed by researchers for its almost obvious association with student achievement by many researchers. And indeed, content coverage is often significantly related to student performance. Yet, this is not the whole story, because implementation of adequate content coverage in education leaves much to be desired (Boscardin et al., 2005). A possible explanation for this is that schools are focussing too much on covering the content by teaching facts instead of helping the students truly understand concepts (Gau, 1997). Apparently, teaching for understanding concepts is not the same as teaching the textbook content. Another explanation for the absence of adequate content coverage may be related to the risk of teaching to the test (Boscardin et al., 2005). When you present teachers the exact test content, teachers can adjust their teaching content. This leads to higher student grades, while students do not learn more or deeper. When gains in student scores are larger than gains in real student learning, this is considered a case of grade inflation (Jennings & Bearak, 2014).

OTL from the perspective of different research traditions

OTL has been examined from the perspective of three different research traditions: educational effectiveness research, curriculum research and achievement test development (Scheerens et al., 2017). They have in common that they all underline the key role of alignment within the concept of OTL. However, all three in a different way. Besides, within these traditions there is discussion about which aspects encompass OTL. Some authors state that OTL includes not only context exposure, but also variables like teaching quality, time on tasks and school resources (Santibanez & Fagioli, 2016). A broad picture of OTL from the three traditions will be outlined, before clarity about the scope of OTL used in this research is given.

The first research tradition, educational effectiveness research, looks at all factors within the educational system that affect student outcomes. The focus is not always on students' academic growth, but also on their social development (Reynolds, Sammons, De Fraine, Townsend, & Van Damme, 2011). OTL is one of the conditions tested in educational effectiveness studies. In these studies, OTL is the independent variable and measures of student outcomes the dependent variable (Scheerens et al., 2017). The purpose of effectiveness research in OTL has always been investigation of the effect of OTL on student achievement.

In the tradition of curriculum research, the concept of alignment is much broader. It might include alignment between the intended and implemented curriculum, or the intended and tested curriculum, without implication of student achievement. Alignment between the intended, implemented and tested curriculum is often seen as a positive OTL.

At last, OTL in the light of achievement test development, in which OTL is seen as an interpretation of test preparation. OTL in this research tradition raises a lot of questions, because it is often associated with teaching to the test. Looking at the similarity between the teaching content and the test content one might wonder if similarities can be too big, which makes the test outcome unreliable or even invalid. A gain in test scores is in these cases not always the result of gains in learning. Sahlberg (2011), states that standardised tests will not lead to improvements in education. Instead, teachers will adjust their teaching to these tests, causing increasing student scores. Teachers adjust their teaching by aligning instruction with standards, emphasizing predictably tested standards and teaching skills in tested formats (Jennings & Bearak, 2014). In the Netherlands, tests from the Central Institute for Test Development (Cito) brought up the same discussion. Every few years they have to develop a new test because results are getting higher every year because the test items become too well-known. A vicious circle is created, because the annually increasing student scores lead to new tests, new tests cause a different way of teaching (content, methods, etc.), and the adapted way of teaching results in higher test scores again (Zwik, 2014). At the same time, there are also positive aspects of teaching to the test. For example, Jennings & Bearaks (2014) put forward that: ''teaching students test-taking skills that are specific to a test form may allow students to more accurately demonstrate their knowledge of the tested skills and content'' (p. 2). But they have counterarguments too, which makes teaching to the test a debatable subject.

OTL in earlier research

Scheerens (2015) concluded that the number of meta-analyses about OTL is limited. Most of the meta-analyses that are available highlight various aspects of OTL, which makes them hard to compare. Kablan (2013) focused on material use in classroom instruction, Kyriakides (2013) on effective teaching, Schroeder (2007) on teaching strategies and Spada & Tosmita (2010) on implicit and explicit teaching. All these topics resemble OTL, but they are not equal to OTL. Scheerens et al. (2017) mentions their study variables as: ''relatively remote proxies of OTL'' (p. 27). A review study that does deal with OTL as it is meant in this study is that of Squires (2012) about the research around curriculum alignment. He concludes that student achievement can be improved by aligning the written, taught and tested curriculum, but he gives no average effect size. A study that does come with an average effect size of OTL, defined as the alignment between taught and tested content, is Creemers (1994). He described an average effect size of d=0.88. This is very high, but only based on four older studies. Besides Creemers, Scheerens (2007) and Hattie (2009) also investigated OTL in the light op earlier research. But they used a definition of OTL that was broader than just alignment between the written, taught and tested curriculum. They came with an average effect size, using Cohen's d, of respectively 0.30 and 0.39. Both effect sizes are considered a small effect size. A recent meta-analysis or review study about OTL in the sense of content coverage was not available until Scheerens et al. did their meta-analysis in 2017: Opportunity to learn, curriculum alignment and test preparation. They had a different approach than their predecessors by not determining an average effect size, but by looking at the proportion of statistically significant positive effects. They found a proportion of statistically significant positive effects of 43.8%, a modest percentage.

Theoretical framework and its applications to the current study

In this study the perspective of educational effectiveness research applies, because the focus is on the effect of OTL on student achievement. Emphasis will be on the alignment between actual teaching and the test content. Actual teaching includes both classroom lessons with a teacher, as well as moments of independent learning with a textbook. Thereby, the quality of these lessons and textbooks is disregarded, because that would request a multi-dimensional measure of instructional quality. Also the time that students spent on a task or in classrooms is ignored. The focus is only on the content that has been covered.

Methodology

Research design
This study, building on Scheerens et al. (2017), is a meta-analysis about OTL including dozens of primary empirical studies. For each study, a distinction is made between significant (with a significance level of 0.05) and non-significant effects and between positive and negative significant effects. After determining if the inclusion criteria from Scheerens et al. (2017) are still appropriate and complete, the studies were re-evaluated. The studies that were found useful were listed a table equal to the one in Scheerens et al. (2017), only without the deleted studies. This table can be found in Appendix 2. Subsequently, a vote-count is carried out in which the proportion of statistically significant positive effects is calculated after categorizing the effects of the studies as non-significant effects, significantly negative effects and significantly positive effects. Finally, this study will examine the influence of certain study characteristics on the percentage of significantly positive effects. The following paragraphs will elaborate on the different steps that were taken.

Identification and collection of studies
The inclusion criteria were similar to the criteria of the original study of Scheerens et al. (2017), namely:

- Studies published between 1995 and 2015.
- Studies with achievement scores as dependent variable.
- Studies executed in primary and secondary regular education.
- Studies reported in Dutch, English or German.
- Studies reporting effect sizes.

Besides these criteria, this study added one more. In the original study, a broader conceptualization of OTL applied than in this thesis. In this study, only the part of OTL concerning content coverage is included. Content coverage includes terms like curriculum coverage, instructional alignment and topic focus.

Search strategy
Studies were gathered by a keyword-based search in the electronic databases of ERIC, PsycARTICLES, Psychology and Behavioral Sciences Collection and PsycINFO. In addition, a backward search was performed in which all reference lists of identified useful articles were scrutinized. When potentially useful articles were found in the reference lists they were further investigated and included in the meta-analysis in case of suitability (Scheerens et al., 2017). More details of the literature search that was conducted are provided in Appendix 1.

Meta-analysis
The literature search from Scheerens et al. (2017) was completed in December 2015. A total of 6006 studies were found and assessed for relevance. Of the 6006 articles, 51 met the inclusion criteria as mentioned in paragraph 3.2. This study uses to the utmost extent the same studies as the original study. However, some changes have been made, due to more strictly adhering to the inclusion criteria, the tightened criteria about content coverage and splitting up an article that described two studies. Hereafter, 38 articles remained for this study. These studies are succinctly described in a table to be found in Appendix 2. The table is similar to the table in Scheerens et al. (2017), only

without the articles that are excluded for this study.

Changes compared to the original study

A few changes were made compared to Scheerens et al. (2017) considering the used articles. The changes and the reasons for these amendments are listed in Appendix 3.

Data-analysis

The quantitative data gathered from the meta-analysis is analysed using IBM's SPSS 24. First, descriptive statistics were computed to get a clear overview of all the collected data. For six study characteristics it will be made insightful how many articles belong to each category. In addition, a vote-count is conducted in which the proportion of significant and non-significant effects is demonstrated. Secondly, multiple chi-square tests and cross tabulations are used to determine if there were differences between groups concerning the six study characteristics. For these tests, multiple effects in one article are not considered as a whole, but all effects are separately dealt with. This to prevent groups being too small for valid statistical tests.

Results

Descriptive statistics

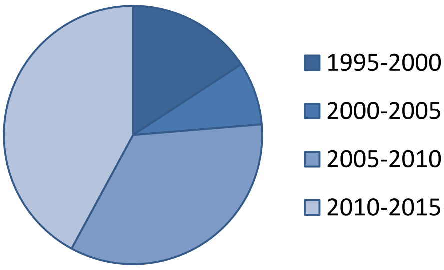
For this study 38 studies have been used. Different study characteristics are examined. The following tables and graph show an overview of these characteristics. The graphs depict the number of articles and effects belonging to a certain category. The dotted lines in the tables provide a subdivision between the number of articles belonging to the categories on the one hand and the number of effects, as used in the chi-square tests, belonging to the categories on the other hand. The graphs are based on the number of articles.

Year of publication

Table 1 and graph 1 show the year of publication of the studies. The last decade there has been an increase of research on OTL.

Table and graph 1*: Year of publication*

	Number of articles	Percent	Cumulative Percent	Number of effects
1995-2000	6	15.8	15.8	10
2000-2005	3	7.9	23.7	45
2005-2010	13	34.2	57.9	38
2010-2015	16	42.1	100	47
Total	38	100		140

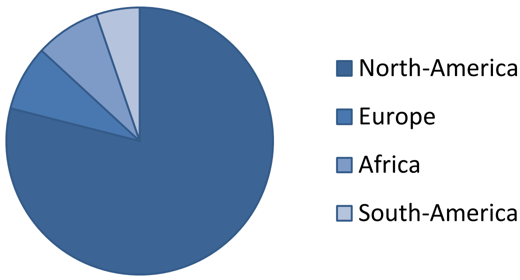


Geographical area

Table 2 and graph 2 give an overview of the geographical area in which the study is performed. The vast majority of the studies is conducted in North-America.

Table and graph 2: *Geographical area*

	Number of articles	Percent	Number of effects
North-America	30	78.9	115
Europe	3	7.9	
Africa	3	7.9	25
South-America	2	5.3	
Total	38	100	140

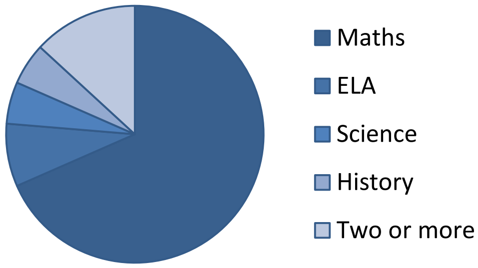


Subject

Table 3 and graph 3 demonstrate the subject that is used to measure student achievement scores. The largest part of the studies was about the influence of OTL on mathematics performance. The category mathematics includes also studies that examined only algebra and the category English Language Arts also includes studies that only examined reading or writing. These subcategories are combined into one category to maintain the overview. Besides, a category that is too small makes some inferential statistics tests impossible.

Table and graph 3: *Subject*

	Number of articles	Percent	Number of effects
Mathematics	26	68.4	116
English Language Arts	3	7.9	
Science	2	5.3	20
History	2	5.3	
Two or more subjects	5	13.2	
Missing			4
Total	38	100	136

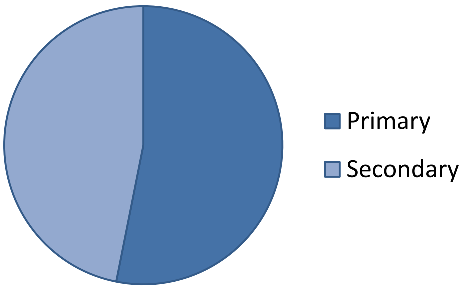


Year of study

Table 4 shows the year of study of the participants. This can be either the students who follow classes from that year of study or the teacher who teaches a class in that year of study. The zero stands for kindergarten. Studies in which eight grade students or teachers are the respondents predominated. Six articles are mentioned as missing in the table and graph since these articles include longitudinal research over several years. They cannot be categorized per year. For the inferential statistics two categories were made for the year of study. The first includes zero up to and including six and the second comprises seven up to and including twelve.

Table and graph 4: *Year of study*

	Number of articles	Percent	Cumulative percent	Number of effects
0	3	7.9	7.9	
1	1	2.6	10.5	
2	2	5.3	15.8	
3	1	2.6	18.4	84
4	2	5.3	23.7	
5	2	5.3	29	
6	6	15.8	44.8	
7	2	5.3	50.1	
8	10	26.3	764	
9	1	2.6	79	49
10	1	2.6	81.6	
12	1	2.6	84.2	
Missing	6	15.8	100	7
Total	38	100		133

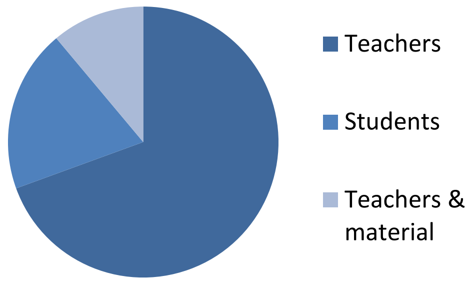


Type of respondents for data collection

Table 5 and graph 5 give an overview of the type of respondents that are used for data collection. Most researchers chose to derive their information on OTL from teachers from interviews, surveys, questionnaires, reports and teacher logs.

Table and graph 5: *Type of respondents for data collection*

	Number of articles	Percent	Number of effects
Teachers	25	65.8	105
Students	7	18.4	16
Teachers & material	4	10.5	13
Missing	2	5.3	6
Total	38	100	134

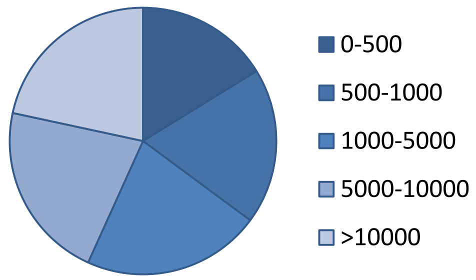


Number of respondents

Table 6 shows the number of respondents that are used to determine the average achievement score of the students.

Table and graph 6: *Number of respondents*

	Number of articles	Percent	Number of effects
0-500	6	15.8	25
500-1000	7	18.4	10
1000-5000	8	21.1	47
5000-10000	8	21.1	17
>100000	8	21.1	32
Missing	1	2.6	9
Total	38	100	131



Vote-count

The effect sizes that are measured in the 38 articles are displayed in a vote-count in table 7. It shows the number of effect sizes, whether they are positive or negative and whether the association was statistically significant at a 5% level.

Table 7: *Vote-count effect sizes*

Article	Number of OTL effects	Number of statistically significant effects (p<.05)	Number of statistically non-significant effects (p<.05)	Number of statistically significant positive effects (p<.05)	Number of statistically significant negative effects (p<.05)
Aguirre-Muñoz & Boscardin (2008)	2	2	0	2	0
Boscardin, Aguirre-Muñoz, Stoker, Kim, Kim & Lee (2010)	1	1	0	1	0
Carnoy & Arends (2012)	4	0	4	0	0
Claessens, Engel & Curran (2012)	12	8	4	5	3
Cogan, Schmidt & Wiley (2001)	2	2	0	2	0
Cueto, Guerrero, Leon, Zapata, & Freire (2014)	3	2	1	2	0

Cueto, Ramirez & Leon (2006)	3	2	1	2	0
D'agostino et al. (2007)	3	2	1	1	1
Desimone, Smith & Phillips (2013)	4	4	0	2	2
Elliott (1998)	4	4	0	2	2
Engel, Claessens & Finch (2013)	4	3	1	2	1
Gamoran et al. (1997)	1	0	1	0	0
Gau (1997)	1	1	0	1	0
Heafner & Fitchett (2015)	2	2	0	1	1
Herman & Abedi (2004)	1	1	0	1	0
Holtzman (2009)	4	1	3	1	0
Kurz, Elliott, Kettler & Nedim (2014)	1	0	1	0	0
Kurz, Elliott, Wehby & Smithson (2010)	8	3	5	3	0
Marsha (2008)	2	0	2	0	0
Mo, Singh & Chang (2013)	1	1	0	1	0
Niemi, Wang, Steinberg, Baker & Wang (2007)	2	1	1	1	0
Oketch, Mutisya, Sagwe, Musyoka & Ngware (2012)	1	0	1	0	0
Ottmar, Konold & Berry (2013)	2	1	1	1	0
Plewis study A (1998)	1	1	0	1	0
Plewis study B (1998)	1	1	0	1	0
Polikoff & Porter (2014)	6	0	6	0	0
Ramirez (2006)	1	1	0	1	0
Reeves (2005)	1	1	0	1	0
Reeves & Major (2012)	1	1	0	1	0
Reeves, Carnoy & Addy (2013)	2	1	1	0	1
Schmidt (2009)	3	1	2	1	0
Schmidt, Cogan, Houang & McKnight (2009)	1	1	0	1	0
Snow-renner (2001)	32	14	18	14	0
Tarr, Ross, McNaught, Chávez, Grouws, Reys, Sears & Taylan (2010)	3	3	0	3	0
Törnroos (2005)	9	2	7	2	0
Wang (1998)	2	2	0	2	0
Wang (2009)	5	4	1	4	0
Wonder-McDowell, Reutzel & Smith (2011)	4	4	0	4	0
Total of studies	Total effects	Total significant effects	Total in-significant effects	Total significant positive	Total significant negative
38	140	78	62	67	11

The results in table 7 show that 78 out of the 140 effects are statistically significant, this is more than half. The most important indicator is the proportion of statistically significant positive effects, which is 47.9% (67/140*100). In the original study, this proportion was 43.8%. The stricter compliance with the criteria and focus only on content coverage, have caused a small increase of relatively more statistically significant positive effects.

Relationship between significant positive effects and study characteristics

Per study characteristic, a chi-square test is conducted to investigate if there are differences between groups concerning the proportion of statically significant positive effects. To use a chi-square test the following two conditions must be met: all the expected frequencies must be met, and up to 20% of the expected frequencies may be smaller than five. These conditions will be checked per study characteristic. Interpretation of the chi-square value is done based on a distribution table (to be found in Appendix 4) with the critical values of chi-square per number of degrees of freedom. A large chi-square value means a greater difference between the actual and the expected data. If the chi-square value is greater than the critical value there is a significant difference.
In addition to the chi-square tests, cross tabulations have been drawn up to facilitate interpretation, by giving an overview of the number and percentage of significant positive effects and the other effects (non-significant and significant negative) per category.

Year of publication

Table 9: *Cross tabulation year of publication – proportion statistically significant positive effects*

Year of publication		Statistically significant positive effects	Non-significant and significantly negative effects
1995 – 2000	Count	7	3
	Expected count	4,8	5,2
	% within year of publication	70%	30%
2000 – 2005	Count	20	25
	Expected count	21,5	23,5
	% within year of publication	44.4%	55.6%
2005 – 2010	Count	21	17
	Expected count	18,2	19,8
	% within year of publication	55.3%	44.7%
2010 - 2015	Count	19	28
	Expected count	22,5	24,5
	% within year of publication	40.4%	59.6%
Total	Count	67	73
	Expected count	67	73
	% within year of publication	47.9%	52.1%

One cell out of eight (12.5%) has an expected count less than 5 (4,8). There was no significant association between the proportion of significant positive effects and the year of publication, $\chi2(3) = 4.050$, $p = .256$. Effects with a publication year between 1995 and 2000 were more likely to have a statistically significant positive effect (70%) than articles published in 2000-2005 (44.4%), 2005-2010 (55.3%) and 2010-2015 (40.4%).

Geographical area

Table 9: *Cross tabulation geographical area – proportion statistically significant positive effects*

Publication area		Statistically significant positive effects	Non-significant and significantly negative effects
North-America	Count	58	57
	Expected count	55	60
	% within geographical area	50.4%	49.6%
Other	Count	9	16
	Expected count	12	13
	% within geographical area	36%	64%
Total	Count	67	73
	Expected count	67	73
	% within geographical area	47.9%	52.1%

All expected cell frequencies were greater than five. There was no statistically significant association, $\chi 2(1) = 1.715$, $p = .190$. Effects in articles that are published in North-America are more likely to be significantly positive (50.4%) than effects in articles that are published in other areas (36%).

Subject

Table 10: *Cross tabulation subject – proportion statistically significant positive effects*

Subject		Statistically significant positive effects	Non-significant and significantly negative effects
Mathematics	Count	52	64
	Expected count	56,3	59,7
	% within subject	44.8%	55.2%
Other	Count	14	6
	Expected count	9,7	10,3
	% within subject	70%	30%
Total	Count	66	70
	Expected count	66	70
	% within subject	48.5%	51.5%

All expected cell frequencies were greater than five. There was a statistically significant association, $\chi 2(1) = 4.327$, $p = .038$. Effects in studies with mathematics as variable are less likely to be significantly positive (44.8%) than effects in studies with other subjects as variable (70%).

Type of education

Table 11: *Cross tabulation type of education – proportion statistically significant positive effects*

Type of education		Statistically significant positive effects	Non-significant and significantly negative effects
Primary education	Count	38	46
	Expected count	41,7	42,3
	% within type of education	45.2%	54.8%
Secondary education	Count	28	21
	Expected count	24,3	24,7
	% within type of education	57.1%	42.9%
Total	Count	66	67
	Expected count	66	67
	% within type of education	49.6%	50.4%

All expected cell frequencies were greater than five. There was no statistically significant association between, $\chi2(1) = 1.754$, $p = .185$. Effects in articles with primary education as variable are less likely to be significantly positive (45.2%) than effects measured in secondary education (57.1%).

Type of respondents for data collection

Table 12: *Cross tabulation data collection – proportion statistically significant positive effects*

Data derived from…		Statistically significant positive effects	Non-significant and significantly negative effects
Teachers	Count	50	55
	Expected count	48,6	56,4
	% within data collection	47.6%	52.4%
Students & student material	Count	6	10
	Expected count	7,4	8,6
	% within data collection	37.5%	62.5%
Teachers & student material	Count	6	7
	Expected count	6	7
	% within data collection	46.2%	53.8%
Total	Count	62	72
	Expected count	62	72
	% within data collection	46.3%	53.7%

All expected cell frequencies were greater than five. There was no statistically significant association, $\chi2(2) =.572$, $p = .751$. The percentage of statistically significant positive effects is a little higher when the data is derived from teachers (47.6%) compared to the group of effects where data is derived from students and student material (37.5%) and from teachers and student material (46.2%).

Number of respondents

Table 13: *Cross tabulations number of respondents – proportion statistically significant positive effects*

Number of respondents		Statistically significant positive effects	Statistically significant effects
0-500	Count	11	14
	Expected count	12,4	12,6
	% within number of respondents	44%	56%
500-1000	Count	6	4
	Expected count	5	5
	% within number of respondents	60%	40%
1000-5000	Count	25	22
	Expected count	23,3	23,7
	% within number of respondents	53.2%	46.8%
5000-10000	Count	7	10
	Expected count	8,4	8,6
	% within number of respondents	41.2%	58,8%
>10000	Count	16	16
	Expected count	15,9	16,1
	% within number of respondents	50%	50%
Total	Count	65	66
	Expected count	65	66
	% within number of respondents	49.6%	50.4%

One cell (10%) has an expected count less than 5 (4,96). There was no statistically significant association, $\chi2(4) = 1.473$, $p = .831$. Effects in studies with 500 to 1000 respondents are most likely to be significantly positive (60%).

Conclusion and discussion

The aim of this study was to examine if an explanation for the modest proportion of statistically significant positive effects of OTL in Scheerens et al. (2017) could be found in the broad concept description of OTL or in the influence of study characteristics.

The first possible explanation, a broad concept description, is tested by a vote-count after re-evaluation of the studies used in Scheerens et al. (2017). An additional inclusion criterion was added, namely that only studies with a focus on content coverage were found appropriate. The vote-count made a distinction between non-significant effects, significant positive effects en significant negative effect. The proportion of statistically significant positive effects increased from 43.8% to 47.9%, after including only content coverage as OTL variable. This is a merely minimal improvement, which does not explain the modest proportion of statistically significant positive effect in Scheerens et al. (2017). However, it should be considered that a vote-count is not the most advanced technique for comparisons. So are for instance all the non-significant positive effects grouped under non-significant effects. No further research has been conducted into the degree of (non)significance or the number of non-significant positive effects. Besides, comparisons of the exact effect sizes per study would be more precise. The reason that this has not been done in this study lies in the fact that it would become to extensive and time-consuming for a master's thesis.

The second possible explanation, the influence of study characteristics, is examined by conducting multiple chi-square tests. Only one out of the six tests showed a significant difference in the proportion of significant positive effects. This was the case with the study characteristic subject, in which effects in studies with mathematics as variable are less likely to be significantly positive (44.8%) than effects in studies with other subjects as variable (70%). The six chi-square tests belonging to the six study characteristics are further described below.

<u>Year of publication</u>: Regarding the year of publication, it was expected that the older the studies the stronger the OTL effect would be. And indeed, the cross tabulation showed that the effects in the oldest articles (1995-2000) were most likely to be significantly positive (70%). However, because the differences were not that big, the chi-square test did not give a significant result ($\chi 2(3) = 4.050$, $p = .256$).

<u>Geographical area</u>: Because grading leniency, due to the great importance that is attached to student evaluations, is a well-known phenomenon in North-America, it is expected that OTL effects in North-America are weaker than in other continents. However, results show that effects in articles that are published in North-America are more likely to be significantly positive (50.4%) than effects in articles that are published in other areas (36%). The difference is relatively small, so the chi-square test gave no significant result ($\chi 2(1) = 1.715$, $p = .190$).

<u>Subject</u>: For the study characteristic subject it was predicted that studies with mathematical achievement as variable would result in stronger OTL effects compared to studies with other subjects as variable. This because mathematical achievement is little influenced by other factors than school. The chi-square test for this study characteristic gave a significant result ($\chi 2(1) = 4.327$, $p = .038$), but not in the expected direction. Effects in studies with mathematics as variable are less likely to be significantly positive (44.8%) than effects in studies with other subjects as variable (70%). Perhaps an explanation can be found in the fact that mathematics is a subject with a fixed content. Worldwide, more or less the same content is offered and tested. With a subject such as history or science that is different, where often more content is offered than tested. When a school is free in which topics to discuss and test, a stronger OTL effect could be found.

<u>Type of education</u>: The expectation of the result of the chi-square test and corresponding cross

tabulation for this study characteristic was uncertain. Primary and secondary education differ in many ways, like the age of the pupils, specialist teachers versus all-round teachers and the opportunities for subject integration. But none of the differences pointed in a clear direction favouring primary or secondary education in terms of the strength of the OTL effect. The chi-square test showed no statistically significant association between the two groups ($\chi2(1) = 1.754$, $p = .185$). However, effects in articles with secondary education as variable have a slightly bigger chance to be significantly positive (57.1%), than effects measured in primary education (45.2%). The difference is small, so no further conclusions can be drawn.

Type of data: A stronger OTL effect was expected for the studies in which the data comes from students. However, this did not appear to be the case. The group of effects were data was derived from students and student materials had the smallest chance to be statistically positive (37.5%) compared to the group with data from teachers and student material (46.2%) and the group with data derived from only teachers (47.6%). The chi-square test revealed that this difference was not statistically significant ($\chi2(2) =.572$, $p = .751$). A possible explanation may lie in the fact that most studies with students as source of information used student's workbooks and notebooks to obtain information. Only a few studies relied on student interviews or questionnaires. Information from student material and students themselves does not have to be the same. For example, a student can forget to mention something in an interview or a student does not write everything down that should be written in his or her notebook.

Number of respondents: For the number of respondents it was expected that there would be more small studies with significant positive OTL effects than large studies by the potential chance of a publication bias. The chi-square test revealed no statistically significant associational ($\chi2(4) = 1.473$, $p = .831$). Studies with 500-1000 respondents are most likely to have significant positive effects (60%) opposite to studies with 5000-10000 respondents which are least likely to have significant positive effects (41.2%). The group with the smallest studies, 0-500 respondents, even have the second lowest chance to have significant positive effects (44%). Therefore, there is no proof of a publication bias in the studies used for this meta-analysis.

The chi-square tests and cross tabulations gave no clear explanation for the relatively disappointing proportion of statistically significant positive OTL effects in Scheerens et al. (2017). The only significant association was found concerning the study characteristic subject. Effects in studies with other subjects than mathematics were more likely to be significantly positive. However, only 20 out of the 136 effects were effects in studies with other subjects than mathematics as a variable. This is just a mere 15 percent. Therefore, more research on other subjects than mathematics is necessary to investigate if the proportion of statistically significant positive OTL effects is really that much higher for other subjects than mathematics.

Another limitation of this study is that a single article can strongly affect a result. For example regarding the study characteristic geographical area. The study of Törnroos (2005) examined nine OTL effects of which seven turned out to be non-significant. The group of effects from articles conducted outside North-America consist of only 25 effects, so 28% of the effects in this group come from Törnroos. With the article of Törnroos included 36% of the effects from articles outside North-America are statistically significant positive, while this percentage raises to 43.8% when this article is excluded.

As a result of this study it can be concluded that both, a narrower concept description of OTL, as well as, the influence of study characteristics, gave no explicit explanation for the modest proportion of statistically significant positive OTL effects in Scheerens et al. (2017). Future research needs to focus on OTL in relationship to achievement in other subjects than mathematics. Besides, a meta-analysis using the exact effect sizes of studies in comparisons would be an interesting addition.

References

Aguirre-Muñoz, Z. & Boscardin, C.K. (2008). Opportunity to learn and English learner achievement: Is increased content exposure beneficial? *Journal of Latinos and Education, 7*(3), 186 -205.

Boscardin, C. K., Aguirre-Muños, Z., Stoker, G., Kim, M., & Lee, J. (2005). Relationship between opportunity to learn and student performance on English and Algebra assessments. *Educational Assessment, 10*(4), 307-332.

Bronneman-Helmers, H.M. (1999). *Scholen onder druk. Op zoek naar de taak van de school in een veranderende samenleving* [Schools under pressure: Looking for the task of schools in a changing society]. Den Haag: Sociaal en Cultureel Planbureau.

Butori, R. & Parguel, B. (2012). When students give biased responses to researchers: An exploration of traditional paper vs. computerized self-administration. *EMAC, 2010*, Copenhague, Denmark.

Carnoy, M. & Arends, F. (2012). Explaining mathematics achievement gains in Botswana and South Africa. *Prospects, 42* (4), 453-468.

Carr, J. (2007). *Approaches to teaching & learning.* Report presented at the INTO consultative conference on education 2007. Dublin: Irish National Teachers' Organization.

Claessens, A., Engel, M. & Curran, F.C. (2014). Academic content, student learning, and the persistence of preschool effects. *American Educational Research Journal, 51*(2), 403-434.

Cogan, L.S., Schmidt, W.H. & Wiley, D.E. (2001). Who takes what math and in which track? Using TIMSS to characterize U.S. students' eighth grade mathematics learning opportunities. *Educational Evaluation and Policy Analysis, 23*(4), 323-341.

Creemers, B. P. M. (1994). *The effective classroom.* London: Cassell.

Crumbley, L., Henry, B.K., & Kratchman, S.H. (2001). Students' perceptions of the evaluation of college teaching. *Quality Assurance in Education, 9*(4), 197-207.

Cueto, S., Ramirez, C. & Leon, J. (2006). Opportunities to learn and achievement in mathematics in a sample of sixth grade students in Lima, Peru. *Educational Studies in Mathematics 62* (1), 25-55.

Cueto, S., Guerrero, G., Leon, J., Zapata, M. & Freire, S. (2014). The relationship between socioeconomic status at age one, opportunities to learn and achievement in mathematics in fourth grade in Peru. *Oxford Review of Education, 40*(1), 50-72.

D'agostino, J., Welsh, M.E. & Nina, M. C. (2007). Instructional sensitivity of a State's Standards-Based Assessment. *Educational Assessment, 12* (1), 1-22.

Desimone, L.M., Smith, T.M. & Phillips, K. (2013). Linking student achievement growth to professional development participation and changes in instruction: A longitudinal study elementary students and teachers in title I schools. *Teachers College Records, 115*(5), 1-46.

Elliott, M. (1998). School finance and opportunities to learn: Does money well spent enhance students' achievement? *Sociology of Education, 71*(3), 223-245.

Engel, M., Claessens, A. & Finch, M. A. (2013). Teaching students what they already know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis, 35*(2), 157-178.

Gamoran, A., Porter, A.C., Smithson, J. & White, P.A. (1997). Upgrading high school mathematics instruction: improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis, 19* (4), 325-338.

Gau, S.-J. (1997). *The distribution and the effects of opportunity to learn on mathematics achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, timeMarch. (ERIC Document Reproduction Service No. ED407231).

Hattie, J. (2009). *Visible learning.* Abingdon: Routledge.

Heafner, T.L. & Fitchett, P.G. (2015). An opportunity to learn US history: What NAEP data suggest regarding the opportunity gap. *The High School Journal, 98* (3), 226-249.

Herman, J. L., Klein, D.C.D., & Abedi, J. (2000). Assessing students' opportunity to learn: Teacher and student perspectives. *Educational Measurement: Issues and Practice, 19*(4), 16-24.

Herman, J.L., & Abedi J. (2004). *Issues in assessing English language learners' opportunity to learn mathematics* (CSE Report No. 633). Los Angeles: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.

Holtzman, D.J. (2009). *Relationships among content standards, instruction, and student achievement.* Retrieved from ProQuest Dissertations and Theses.

Jennings, J.L., & Bearak, J.M. (2014). ''Teaching to the test'' in the NCLB Era: How test predictability affects our understanding of student performance. *Educational Researcher, 43*(8), 381-389.

Kablan, Z., Topan, B., & Erkan, B. (2013). The effectiveness level of material use in classroom instruction: A meta-analysis study. *Educational Sciences: Theory and Practice, 13*(3), 1638-1644.

Kurz, A., Elliott, S.N., Wehby, J.H. & Smithson, J.L. (2010). Alignment of the intended, planned, and enacted curriculum in general and special education and its relation to student achievement. *The Journal of Special Education, 44*(3), 131-145.

Kurz, A., Elliott, S. N., Kettler, R. J., & Yel, N. (2014). Assessing students' opportunity to learn the intended curriculum using an online teacher log: Initial validity evidence. *Educational Assessment, 19*(3), 159-184.

Kyriakides, L., Christoforou, C., & Charalambous, C.Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher education, 36,* 143-152.

Marsha, I. (2008). Using instructional sensitivity and instructional opportunities to interpret students' mathematics performance. *Journal of Educational Research and Policy Studies, 8*(1), 23-43.

Marzano, R. J. (2003). *What works in schools. Translating research into action.* Alexandria, V A: Association for Supervision and Curriculum Development.

Mo, Y., Singh, K. & Chang, M. (2013). Opportunity to learn and student engagement A HLM study on eighth grade science achievement. *Educational Research for Policy and Practice, 12*(1), 3-19.

Niemi, D., Wang, J., Steinberg, D.H., Baker, E.L. & Wang, H. (2007). Instructional sensitivity of a complex language arts performance assessment. *Educational Assessment, 12*(3&4), 215-237.

Oketch, M., Mutisya, M., Sagwe, J., Musyoka, P. & Ngware, M.W. (2012). The effect of active teaching and subject content coverage on students' achievement: Evidence from primary schools in Kenya. *London Review of Education, 10*(1), 19-33.

Ottmar, E.R., Grissmer, D.W., Konold, T.R., Cameron, C.E. & Berry, R.Q. (2013). Increasing equity and achievement in fifth grade mathematics: The contribution of content exposure. *School Science and Mathematics, 133*(7), 345-355.

Perryman, J., Ball, S., Maguire, M. & Braun, A. (2011). Life in the pressure cooker – School league tables and English and Mathematics teachers' responses to accountability in a result-driven era. *British Journal of Educational Studies, 59*(2), 179-195.

Plewis, I. (1998). Curriculum coverage and classroom grouping as explanations of between teacher differences in pupils' mathematics progress. *Educational Research and Evaluation, 4*(2), 97-107.

Polikoff, M.S. & Porter, A.C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis, 20,* 1-18.

Porter, A.C, Smithson, J., Blank, R., & Zeidner, T. (2007). Alignment as a teacher variable. *Applied Measurement in Education, 20*(1), 27-51.

Ramírez, M.-J. (2006). Understanding the low mathematics achievement of Chilean students: A cross-national analysis using TIMSS data. *International Journal of Educational Research, 45*(3), 102-116.

Reeves, E.B. (2012). The effects of opportunity to learn, family socioeconomic status, and friends on the rural math achievement gap in high school. *American Behavioral Scientist, 56*(7), 887-907.

Reeves, C., Carnoy, M. & Addy, N. (2013). Comparing opportunity to learn and student achievement gains in southern African primary schools: A new approach. *International Journal of Educational*

Development, 33(5), 426-435.

Reynolds, D., Sammons, P., Fraine, B. de., Townsend, T., & Damme, J. van. (2011). *Educational Effectiveness Research (EER): A state of the Art Review.* Paper presented at the annual meeting of the International Congress for School Effectiveness and Improvement, Cyprus.

Sahlberg, P. (2011). Finnish lessons, wat Nederland kan leren van het Finse onderwijs [Finnish lessons, what the Netherlands can learn from Finnish education]. Helmond: Onderwijs maak je samen.

Santibañez, L. & Fagioli, L. (2016). Nothing succeeds like success? Equity, student outcomes, and opportunity to learn in high- and middle-income countries. *International Journal of Behavioral Development, 40*(6), 517-525.

Scheerens, J. (2015). *Overzichtstudie naar gelegenheid tot leren* [Overview study on opportunity to learn]. Aanvraagformulier voor een onderzoeksvoorstel Nationaal Regieorgaan Onderwijsonderzoek (NRO). Utrecht: Auteur.

Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness.* Enschede: Universitiy of Twente, Department of Educational Organisation and Management.

Scheerens, J., Lamain, M., Luyten, H & Noort, P. (2017). *Opportunity to Learn, Curriculum Alignment and Test Preparation. A Research Review.* Houten: Springer.

Schmidt, W.H., Cogan, L.S., Houang, R.T. & McKnight, C. (2009). *Equality of educational opportunity: A myth or reality in U.S. schooling.* Lansing, MI: The Education Policy Center at Michigan State University.

Schmidt, W.H. (2009). *Exploring the relationship between content coverage and achievement: Unpacking the meaning of tracking in eighth grade mathematics.* Education Policy Center, East Lansing, MI: Michigan State University.

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, Tse-Yang, & Lee, Yi-Hsuan. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science of Teaching. 44*(10), 1436-1460.

Schwarzer, G., Carpenter, J.R. & Rücker, G. (2015). Small-study effects in meta-analysis. In: Meta-Analysis with R (pp. 107-141). Switzerland: Springer International Publishing.

Snow-Renner, R. (2001). What is the promise of large-scale classroom practice measures for informing us about equity in student opportunities-to-learn? An example using the Colorado TIMSS. Paper presented at the Annual Meeting of the American Educational Research Association. Seattle, WA 10-14 April 2001.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature

A meta-analysis. *Language Learning, 60*(2), 263-308.

Squires, D. (2012). Curriculum alignment research suggests that alignment can improve student achievement. *The Clearing House, 85*(4), 129-135.

Stevens, F. I. (1993). Applying an opportunity-to-learn conceptual framework to the investigation of the effects of teaching practices via secondary analyses of multiple-case study summary data. *Journal of Negro Education, 62*, 232-248.

Stevens, F.I. (1997). *Opportunity to learn science: Connecting research knowledge to classroom practices.* Philadelphia: Mid-Atlantic Lab for Student Success.

Stevens, F. I., Wiltz, L. & Mona, B. (1998). *Teachers' evaluations of the sustainability of opportunity to learn (OTL) assessment strategies. A national survey of classroom teachers in large urban school districts*. Washington, DC.

Stroebe, W. (2016). Why good teaching evaluations may reward bad teaching. *Perspectives on Psychological Science, 11*(6), 800-816.

Tarr, J. E., Ross, D.J., McNaught, M.D., Chávez, O., Grouws, D.A., Reys, R.E., Sears, R. & Taylan, R.D. (2010). *Identification of student- and teacher-level variables in modelling variation of mathematics achievement data.* Paper presented at the Annual Meeting of the American Educational Research Association, Denver, CO.

The United Nations. (1989). Convention on the Rights of the Child. *Treaty Series, 1577,* 3.

Törnroos, J. (2005). Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation, 31*(4), 315-327.

Verhage, H. & De Lange, J. (1996). *Mathematics education and assessment.* Paper presented at the Amesa conference, Freudenthal Institute, The Netherlands.

Wang, A.H. (2009). Optimizing early mathematics experiences for children from low-income families: A study on opportunity to learn mathematics. *Early Childhood Education Journal, 37*(4), 295-302.

Wang, J. (1998). Opportunity to learn: The impacts and policy implications. *Educational Evaluation and Policy Analysis, 20*(3), 137-156.

Wonder-McDowell, C., Reutzel, D. R. & Smith, J.A. (2011). Does instructional alignment matter? Effects on struggling second graders' reading achievement. *The Elementary School Journal, 112*(2), 259-279.

Zeichner, K. (2010). Competition, economic rationalization, increased surveillance, and attacks on diversity: Neo-liberalism and the transformation of teacher education in the U.S. *Teaching and Teacher Education, 26*, 1544 – 1552.

Zwik, M. (2014). De vicieuze cirkel van ‘teaching to the test [The vicious circle of ‘teaching to the test’]’. Geraadpleegd op 18-10-2017, van https://wij-leren.nl/teaching-to-the-test-cito.php

Appendices

Appendix 1: descriptors used in the literature search

Database: ERIC, PsycARTICLES, Psychology and Behavioral Sciences Collection, PsycINFO
Publication date: 1995-2015

''opportunity to learn'' OR ''curricul* align*'' OR ''learn* what is expected'' OR ''access to instruction'' OR ''curricul* exposure'' OR ''test preparat*'' OR ''exam* preparat*'' OR ''instruction* align*'' OR ''instructional sensitivity'' OR ''enacted curricul*'' OR ''curricul* cover*'' OR ''content cover*'' OR ''curricul* implement*''OR ''curriculum teaching'' OR ''curricul* differen*'' OR ''curricul* coherence'' OR ''topic cover*''
AND
''Effectiveness'' OR '' achievement'' OR ''outcome'' OR ''success'' OR ''influence'' OR ''added-value'' OR ''grade''
NOT: ICT
NOT: disab* OR disadvantage*
NOT: material*
NOT: higher education
NOT: business
NOT: special

(Scheerens et al., 2017)

Appendix 2: Overview articles used for meta-analyses

Aguirre-Muñoz & Boscardin (2008)	Focus	This investigation examined the impact of opportunity to learn content and skills targeted by a writing assessment on the achievement of English learners (ELs), including the potential for differential impact of increased exposure to literary analysis and writing instruction (p. 186). This study is characterized as correlational research and is conducted in the USA.
	OTL measure	The six OTL constructs measured by the teacher survey that were included in the validity analysis were expertise, literary analysis content coverage, writing content coverage, classroom processes, assessment practices, and LAPA (Language Arts Performance Assignment) preparation (p.191).
	Respondents	Teachers (N=27) & 6^{th} grade students (N=1.038)
	Dependent variable	Language Arts Performance Assessment score
	Effect size	The ordinal logistic hierarchical linear modelling analyses indicated that **of the nine OTL variables, only two showed significant effects on student performance**: literacy analysis coverage (β= 0.46, p = .03) and writing coverage (β = 0.54, p = .02) (p. 196/197).
	Comments	Results controlled for all of the predictors at the grand mean level, including gender and language proficiency status (p. 196).
Boscardin, Aguirre-Muñoz, Stoker, Kim, Kim & Lee (2010)	Focus	Examination of the impact of OTL variables on student performance on English and algebra assessments. This study also showed that content coverage was positively correlated with student performance in English and algebra (p. 307). This is a correlational study conducted in the USA.
	OTL measure	The findings were based on the Teacher OTL Survey developed by the UCLA National Center for Research on Evaluation, Standards, and Student Testing (CRESST) in collaboration with content experts. The survey had five major sections corresponding to critical aspects of OTL: (a) teaching experience, (b) teacher expertise in content topics, (c) topic coverage, (d) classroom activities, and (e) assessment strategies and preparation.
	Respondents	Students (N=4.715 English & N=4.724 Algebra) & Teachers (N=118 English & N= 124 Algebra)
	Dependent variable	English and Algebra achievement
	Effect size	**Initial analyses revealed that of the five OTL variables, only two were significant predicators**. The other variables were not considered in the final model. The proportion of variance accounted for in the final model by the OTL variables (teacher expertise in content topics and topic coverage) was 0.28 for the algebra test and 0.35 for the English test. This means that including measures of teacher expertise, content coverage, and mean student SES reduced the variance by 28% for the algebra test and 35% for the English test (p. 324).
	Comments	Results controlled for differences in the average SES of the students and individual differences among students, like initial course grades.
Carnoy & Arends (2012)	Focus	The purpose of the study is to test whether and how classroom and school factors contribute to student gains in mathematics learning. From a classroom perspective, the emphasis is on teacher mathematics knowledge, classroom pedagogy and opportunity to learn in of sample of grade 6 classrooms (p. 453). This is a correlational study conducted in South Africa and Botswana.
	OTL measure	Total lessons on topic and total topics taught, measured by analysing the contents of the three best student notebooks each classroom.
	Respondents	6^{th} grade students (N=5.500) & teachers (N= 126)
	Dependent variable	Mathematics achievement

	Effect size	The OTL measures are both significantly related to student learning gains in South Africa with a significance level of p<.10 instead of p<.05 (**total lessons on topic: .0012, p<.10; total topics taught: .0021, p<.10**), but not in Botswana **(total lessons on topic: .0000, p>.10; total topics taught: .0002, p>.10**) (p. 465) .
	Comments	Results are controlled for initial student achievement, several teacher quality variables, characteristics of individual students and students' families, average classroom SES, observed class size and school conditions.
Claessens, Engel & Curran (2014)	Focus	Examination of how reading and mathematics content coverage in kindergarten is associated with the maintenance of preschool skills advantages. Results suggest that increased exposure to advanced content could help maintain preschool skill advantages while promoting the skills of children who did not attend preschool (p.1) This is a correlational study conducted in the USA.
	OTL measure	Content exposure: the influence of four measures of kindergarten academic content – basic mathematics, advanced mathematics, basic reading and advanced reading – on math and reading achievement test scores. Teachers were surveyed about classroom activities and content. Distinction is being made between children who attended Center Care, children who went to a funded childhood program like Head Start and children with Other Care.
	Respondents	Kindergarten students (N=17.981) & teachers (N=3.038)
	Dependent variable	Mathematics and reading achievement
	Effect size	Coefficients from regressions predicting spring mathematics/reading achievement with mathematics/reading content measures by children's preschool experience: <u>Center Care</u>: Basic math (-0.0463, p<0.01), Advanced math (0.066, p<0.01), Basic reading (-0.0192, p<0.05) & Advanced reading (0.0611, p<0.01) (**4/4 significant**). <u>Head Start</u>: Basic math (-0.0215, p>0.1), Advanced math (0.0416, p<0.01), Basic reading (0.00424, p>0.1) & Advanced reading (0.0251, p<0.1) (**2/4 significant**). <u>Other Care</u>: Basic math (-0.0446, p<0.01), Advanced math (0.0531, p<0.01), Basic reading (-0.0086, p>0.1) & Advanced reading (0.0530, p<0.01) (**3/4 significant**) (p.41-42)
	Comments	The OTL measure is based on specific content categories. Analyses control for observable characteristics of teachers and classrooms, a variety of child characteristics and home environment factors that might be correlated with both content measures and student initial achievement (p.15 - 17). Effect net of co-variables and other independent variables.
Cogan, Schmidt & Wiley (2001)	Focus	This article examines the range of eighth-grade mathematics learning opportunities in the U.S. drawing on data gathered for the Third International Mathematics and Science Study (TIMSS). Comparison of students' learning opportunities includes consideration of the specific course in which they were enrolled, the type of textbook employed for the course, and the proportion of time teachers devoted to teaching specific topics (p. 1). This is a correlational study conducted in the USA.
	OTL measure	Variation in 8th grade mathematics in the USA: the opportunity students get to study mathematics. The OTL variables are topic and course-text difficulty. Their influence on mathematics score is being measured by students surveys and teacher questionnaires.
	Respondents	8th grade students (N= over 13.000) and their teachers
	Dependent variable	Mathematics achievement
	Effect size	**Class' topic difficulty and course-text challenge are both significant predictors explaining nearly 40% of the variance in mathematics score across classrooms. Topic coverage has a coefficient of 23.2 with p<0.001 and Course Challenge 13.8 with p <0.001.** Classes exposed to more challenging topics tended to have higher TIMSS scores – on average, 23 points higher for every year increase in the class' international topic

		difficulty. Each increase in a class's course-text challenge rank was associated with nearly a 14 points increase on TIMSS score (p. 20).
	Comments	Results controlled for school's location (urban, rural or suburban), size and percent of minority enrolment. Results are not controlled for initial achievement scores because the study was cross-sectional.
Cueto, Guerrero, Leon, Zapata, & Freire (2014)	Focus	This paper explores the relationship between SES measured at age one, OTL and achievement in mathematics ten years later. (p. 50). This is a correlational study conducted in Peru.
	OTL measure	Four OTL variables were measured: hours of class per year, curriculum coverage, quality of teachers' feedback and level of cognitive demands. Each variable was measured based on exercises found in the notebooks and workbooks, except hours of class per year, which was reported by the head teacher (p.53).
	Respondents	4^{th} grade students (N=104)
	Dependent variable	Mathematics achievement
	Effect size	Curriculum coverage (.44) and the level of cognitive demand (.0.29) indicate a positive and significant association with achievement in mathematics. **With several covariates added, only curriculum coverage remained as a significant predictor of achievement.**
	Comments	Net effect of each variable, results controlled for several covariates, like students' prior abilities.
Cueto, Ramirez & Leon (2006)	Focus	Opportunities to learn mathematics of sixth grade students from 22 public schools in Lima, Peru. Where OTL is defined as curriculum coverage, cognitive demand of the tasks posed to the students, percent of mathematical exercises that were correct and quality of feedback. OTL is positively associated with achievement (p. 25). This is a correlational study conducted in Peru.
	OTL measure	Curriculum coverage, cognitive demand of the tasks posed to the students, percent of mathematical exercises that were correct and quality of feedback. These variables were coded in the workbooks and notebooks of the students, which were gathered at the end of the school year (p. 25).
	Respondents	Sixth grade students (N=369)
	Dependent variable	Mathematics achievement
	Effect size	**Two of the three variables are positive and significant related to achievement**, namely cognitive demands and adequate feedback. When the three OTL variables were included in a factor analysis with varimax rotation one factor resulted, which accounted for 68% of the total variance (p.43).
	Comments	Different independent variables next to the OTL variables, i.e. gender, age, attendance rate during the school year, whether math is the preferred subject for the student, number of persons living at home, the SES score and type of school. Results are not controlled for students' initial achievement level.
D'agostino, Welsh & Corson (2007)	Focus	The purpose of this study was to examine the instructional sensitivity of Arizona's fifth-grade mathematics standards-based assessment (p. 6). For this study, we developed a new method for capturing the alignment between how teachers bring standards to life in their classrooms and how the standards are defined on a test (p. 1). This study is characterized as correlational research and is conducted in the USA.
	OTL measure	Two curriculum experts judged the alignment between how teachers brought the objectives to life in their classrooms and how the objectives were operationalized on the state test (p. 1). Achievement was measured by the fifth-grade mathematics AIMS (Arizona Instrument to Measure Standards). The AIMS math test was designed to measure the Arizona Academic Mathematics standards for Grades 3 through 5. At that time, the standards consisted of six strands: (a) number sense, (b) data analysis and

		probability, (c) patterns, algebra, and functions, (d) geometry, (e) measurement and discrete mathematics, and (f) mathematical structure/logic (p. 9).
	Respondents	5^{th} grade students (N=1.003) and teachers (N=52).
	Dependent variable	Mathematics achievement
	Effect size	To interpret the magnitude of the effect, one can consider a teacher who is one standard deviation above the mean on Alignment and on the Emphasis × Alignment interaction variable. On average, students in the teacher's classroom would be expected to score about 11 scale score points (5.17 points for Alignment **(p<.05)** and 5.75 points for the interaction **(p<.05)**) higher than students, on average, in classrooms at the grand mean of both predictors, which was about a one-fifth standard deviation difference (from Table 1, the standard deviation for the outcome was 55.46). Notice that Emphasis alone has a negative coefficient, -2.00 **(p>.05)** (p. 17).
	Comments	Results controlled for initial achievement differences between classrooms and student socioeconomic status.
Desimone, Smith & Phillips (2013)	Focus	This study examines relationships between teachers' participation in professional development and changes in instruction, and between instruction and student achievement growth (p. 4). This is a correlational study conducted in the USA.
	OTL measure	Minutes per day spent on mathematics, topic focus (basic math and advanced math) and cognitive demands (memorize facts and solve novel problems) are the five OTL variables which are related to the initial status of students achievement and to achievement growth. Reports of professional development and instruction (time spent on mathematics instruction, topic focus, type of learning required or cognitive demands) are taken from teacher's self-report surveys. Student achievement was measured by a special administration of a set of open-ended questions from the Stanford Achievement Test, Ninth Edition, which assessed problem solving and procedures (p. 6).
	Respondents	3th - 5^{th} grade students (N=4.803) and teachers (N=457)
	Dependent variable	Mathematics achievement
	Effect size	Minutes per day spent on mathematics did not significantly predict either initial achievement status or growth. Increased emphasis on Memorizing Facts was associated with slower than average growth in achievement (b=-6.02, p<.037). Emphasis on Solving Novel Problems was associated with extremely modest achievement growth (b=.69, p<0.041) (p.30, 31). The correlation between Focus on Basic Math Topics and achievement growth is -0.042, with p=0.036. For Focus on Advanced Math Topics b=0.061, with p=0.043 (p.55). **Looking at the initial status only 1 out of the 5 variables is significant (Focus on advanced math topics). When it comes to growth 4 out of 5 are significant of which two are positively and two are negatively related.**
	Comments	Results controlled for teacher, school and student characteristics, like teacher's years of experience, school enrolment and initial achievement level.
Elliott (1998)	Focus	This article illuminates both the relationship between spending practices and students' achievement and the specific components of OTL in classroom that affect students' outcomes. Moreover, it indicates how financial resources indirectly affect students' achievement by creating differential access to OTL (p.223). This is a correlational study conducted in the USA.
	OTL measure	Key links between expenditures and achievement: the effect of expenditures on teachers' effectiveness and the effect of expenditures on classroom resources (p.226). Achievement was measured by the 10^{th} grade IRT theta scores, a mathematical

		transformation of the standardized test score is designed to reflect over time.
	Respondents	8^{th} – 10^{th} grade students (N=14.868)
	Dependent variable	Mathematics and science achievement
	Effect size	Expenditures correlate significantly with most measures of OTL. 8 out of 9 OTL variables correlate significantly with expenditures for math students (ranging from -.126 to .142), and 6 out of 7 OTL variables correlate significantly with expenditures for science students (ranging from -.139 to .191). **9 out of 9** OTL variables correlate significantly with students' IRT math test score (ranging from -.059 to .330) and **6 out of 7** OTL variables correlate significantly with students' IRT science test score (ranging from -.066 to .186).
	Comments	Results controlled for student background characteristics (e.g. SES, racial background and gender) and school characteristics. The 8^{th}-grade IRT theta score was controlled in all analysis such that the true outcome was actually gains in math or science achievement between the 8^{th} and 10^{th} grade (p.229).
Engel, Claessens & Finch (2013)	Focus	This study explored the relationship between students' school-entry math skills, classroom content coverage, and end-of-kindergarten math achievement (p.157). This is a correlational study conducted in the USA.
	OTL measure	Exposure to specific mathematics content (the OTL variables: basic counting and shapes, patterns and measurement, place value and currency, and addition and subtraction) and children's early math skills, measured by teacher reports.
	Respondents	Students in kindergarten (N=11.517) & teachers (N=2.176)
	Dependent variable	Mathematics achievement
	Effect size	Devoting additional days per month to Basic Counting and Shapes was negatively associated with the end-of-kindergarten mathematics test scores (-.02 SD). For Patterns and Measurement there was no statistically significant association. For Place Value and Currency there was an increase (.03 SD), and for Addition and Substraction as well (.04 SD). **3 out of 4 OTL variables are significantly correlated to student achievement.**
	Comments	Effect net of co-variables and other independent variables. Results are controlled for initial reading and math skills and cognitive ability (p. 164).
Gamoran, Porter, Smithson, & White (1997)	Focus	In this article, the authors evaluate the success of ''transition'' math courses in California and New York, which are designed to bridge the gap between elementary and college-preparatory mathematics and to provide access to more challenging and meaningful mathematics for students who enter high school with poor skills (p.325). This study is conducted in the USA.
	OTL measure	The extent to which mathematical content and cognitive demands were included in sample classes. Content coverage reflects both the proportion of instructional time that was spent covering tested content and the match of relative emphases of types of content between instruction and the test (19 content areas). Teacher questionnaires provided information on the extent to which the topics on our tests were covered in the sample classes and whether the cognitive demands made on the test also occurred in mathematics instruction (p.330).
	Respondents	7^{th} grade students (N=882)
	Dependent variable	Mathematics achievement
	Effect size	More rigorous content coverage accounts for much of the achievement advantage of college-preparatory classes (p.325). **The correlation between content coverage and instructional effects is 11.615 with p<.10** (p.334)
	Comments	Results are controlled for prior achievement and other student characteristics.
Gau (1997)	Focus	The focus of this paper is further understanding of the distribution and the effects of an

		expanded conception of OTL on student mathematics achievement. In addition to descriptive statistics, a set of two-level hierarchical linear models was employed to analyse a subset of the restricted-use National Education Longitudinal Study of 1988 database. The results revealed that on different scales, various kinds of opportunities to learn mathematics are associated with student mathematics achievement, and opportunities are unequally distributed among different categories of schools (p. 3). This is a correlational study conducted in the USA.
	OTL measure	Content and level of instruction (high achievement group, textbook coverage, instructional time and weekly homework). The variables are measured by students' surveys and teachers' questionnaires. This study cites resources and teachers' mathematical knowledge also as OTL variables.
	Respondents	8^{th} grade students (N=9.702) and their teachers
	Dependent variable	Mathematics achievement
	Effect size	The results of the content and level of instruction analyses are mixed. **Three of the four OTL variables are statistically significant in a positive direction, while the other is significant but negative** (p. 15). The effects of teachers' mathematical knowledge are significant, but the effects of school mathematical resources are not significant.
	Comments	Results controlled for teachers' mathematical knowledge, content and level of instruction, school mathematical resources, gender, race, SES, prior achievement, school sector, minority concentration, community type and school average student SES (p. 3, 4).
Heafner & Fitchett (2015)	Focus	The authors examine National Assessment of Educational Progress in U.S. History (NAEP-USH) assessment data in order to better understand the relationship between classroom- and student-level variables associated with historical knowledge as measured in the 12th grade. Findings document that instructional exposure (OTL) is a factor associated with learning outcomes (p. 226). This is a correlational study conducted in the USA.
	OTL measure	Two categories of instructional exposure: Multimodel Instruction (based on work on group project, give presentation to the class, write a report, use books or computers in library for schoolwork, listen to information presented online, go on field trips or have outside speakers and watch movies or videos) and Text-Dependent Instruction (based on frequency of report writing, discussing material studied, reading extra material, read material from textbook, use publications of historical people, write short answer to questions). OTL variables measured by student surveys.
	Respondents	12^{th} grade students (N=8.610)
	Dependent variable	History achievement
	Effect size	Analysis of the exposure to instruction factors indicated that for each standard deviation increase in text-dependent instruction, NAEP-USH scores increased by 8,61 points (**p<.001**, SE 0.38) where the mean is 250. Conversely, each standard deviation increase in exposure to multimodel instruction was associated with a decrease of 7.48 (**p<.001**, SE 0.49) (p. 236/237).
	Comments	The OTL measure is rather global, not based on more specific content categories. Effect net of co-variables and other independent variables.
Herman & Abedi (2004)	Focus	Exploration of two complimentary approaches for exploring English Language Learners' (ELL) opportunity to learn Algebra 1, representing opposite ends of the cost continuum (p. 6). This is a correlational study conducted in the USA.
	OTL measure	Content coverage measured by surveys of teachers and student, 28 content areas are listed. Teacher-student interactions details through observation.

	Respondents	Survey study:8th grade students (N=602) and teachers (N=9). Observation phase: nine classes of students (N=271) and their teachers.
	Dependent variable	Algebra achievement
	Effect size	Results suggest that OTL is a more determining factor in algebra achievement for ELL students than for the non-ELL group (p. 13). **Results show that the classroom-level OTL measure has significant effects on the outcome variable. However, after accounting for the classroom-level OTL measure, the student-level preparation/OTL factor had no significant effect** (p.16).
	Comments	Three multiple regression models, for all students, for ELLs and for non-ELLs. Each model is controlled for prior math ability and prior student preparation (p. 13).
Holtzman (2009)	Focus	This dissertation addresses the following questions: (1) To what extent is the content of instruction aligned with the California content standards and with the blueprint for the California Standards Test (CST)? (2) How do instruction, the standards, and the CST blueprint compare with one another in the topics covered and the levels of cognitive demand emphasized? (3) To what extent is the alignment of instruction with either the standards or the CST blueprint related to student achievement on the CST? (p. iv). The last question is addressed separately for each school-level (grades 3-6 or grades 6-8) and subject-area (ELA or maths) combinations. This is a correlational study conducted in the USA.
	OTL measure	Topic coverage and cognitive demand emphases in classroom instruction; The data were from a survey of middle school teachers in San Diego City Schools (SDCS). The survey presents teachers with a list of highly detailed topics. For each of the specific topics, teachers first fill in the amount of time spent on the topic by their class during the past school year, and then indicate the proportion of the total time spent on the topic designed to help students meet expectations in each of five different categories of cognitive demand. Student achievement data were provided by SDCS. Scaled scores on the CST in ELA and math for years 2002-03, 2003-04, and 2004-05 are used (p. 41, 42). OTL variables: 1) Alignment with Standards: Overall, 2) Alignment with Standards: Topic, 3) Alignment with Standards: Cognitive Demand, 4) Alignment with CST Blueprint: Overall, 5) Alignment with Blueprint: Topic and 6) Alignment with CST blueprint: Cognitive Demand.
	Respondents	Teachers (N=724), ELA students grades 3-6 (N=2715), Math students grades 3-6 (N=2946), ELA students grades 6-8 (N=1753), and Math students grades 6-8 (N=2556).
	Dependent variable	English language arts (ELA) and mathematics achievement
	Effect size	<u>Elementary ELA results</u>: **All the correlations are negative and not statistically significant** (p<.05). <u>Elementary Math results</u>: 4 out of 6 correlations are negative and not significant (p<.05). **Both positive correlations are statistically significant.** <u>Middle School ELA results</u>: 4 out of 6 variables are positive of which **3 are statistically significant** (p<.05). **One of the 2 negative correlations is significant**. <u>Middle School Math results</u>: 4 out of 6 correlations are positive of which **only one is statistically significant**. The other 2 correlations are negative and not statistically significant (p<.05) (p. 138).
	Comments	Results controlled for student prior achievement, student demographics, and teacher characteristics.
Kurz, Elliott, Kettler & Nedim (2014)	Focus	This study provides initial evidence supporting intended score interpretations for the purpose of assessing OTL via an online teacher log. MyiLOGS yields 5 scores related to instructional time, content and quality. Agreements between log data from teachers and independent observers were comparable to agreements reported in similar studies.

		Moreover, several OTL scores exhibited moderate correlations with achievement and virtually nonexistent correlations with a curricular alignment index (p. 159). This is a correlational study conducted in the USA.
	OTL measure	The extent to which a teacher dedicates instructional time to cover the content prescribed by intended standards using a range of cognitive processes, instructional practices and grouping formats. MyiLOGS scores are designed to allow interpretations about time spent on academic standards, content coverage of academic standards, emphases along a range of cognitive processes, emphases along a range of instructional practices and emphases along a range of instructional grouping formats (p. 165, 166). Each teacher received the standard professional development on the use of MyiLOGS and each teacher participant was observed at least once during his or her logging period (p. 171).
	Respondents	General and special education teachers (N=38) & 8th grade students (N=56).
	Dependent variable	Mathematics and reading achievement
	Effect size	Three out of five OTL variables are significantly related to average class achievement. The correlation between the yearly summary score for Time on Standards and class achievement was **r=.56, p < .05**, accounting for about 31% of the variance in average class achievement. The correlation between the yearly summary score for Cognitive Processes and class achievement was **r= .64, p < .05**, accounting for about 41% of the variance in average class achievement. Last, the correlation between the yearly summary score for Grouping Formats and class achievement was **r= .71, p < .05**, accounting for about 50% of the variance in average class achievement (p.177).
	Comments	Results controlled for state and subject and not for students' prior achievement.
Kurz, Elliott, Wehby & Smithson (2010)	Focus	Examination of the content of the planned and enacted eighth-grade mathematics curriculum for 18 general and special education teachers and the curricula's alignment to state standards via the Surveys of the Enacted Curriculum (SEC). The relation between alignment and student achievement was analyzed for three formative assessments and the corresponding state test within a school year (p. 131). This is a correlational study conducted in the USA.
	OTL measure	Measurement of students' OTL the enacted curriculum and qualification of the alignment of the enacted curriculum to state standards by using SEC as traditional end of year surveys. In addition the surveys were administered midyear to allow for reporting across a shorter period of time. To supplement the standard use of the SEC, the SEC was employed as a prospective survey to measure teachers' planned curriculum at the beginning of the school year. Last, the SEC's alignment statistics were used to examine the presume relation between alignment and achievement (p. 134).
	Respondents	8th grade students (N=238) & teachers (N=18)
	Dependent variable	Mathematics achievement
	Effect size	Significant correlations between student achievement averages and teacher alignment indices were equal to or greater than .48 **(significant 10 out of 15)**. When teacher groups were examined separately, the relation between alignment and achievement remained significant only for special education, with correlations equal to or greater than .75 (p. 131).
	Comments	Results not controlled for other independent variables or co variables (including prior achievement). Only the distinction between special and regular education is being made.
Marsha (2008)	Focus	This exploration includes multiple measures of classroom instruction to evaluate the instructional sensitivity of multiple measures of math achievement and applies an analytic method that makes it possible to relate student-level outcomes to teacher-level

		measures of instruction (p. 23). This is a correlational study conducted in the USA.
	OTL measure	Instructional sensitivity, a link between instructional opportunities and performance on particular assessment items, by measuring two different performance levels, proximal and distal, with students assessments, teacher assessments and teacher interviews.
	Respondents	Third grade students (N= 486) & third grade teachers (N=24)
	Dependent variable	Mathematics and algebraic reasoning
	Effect size	The correlation between prior student achievement and the outcome measure was highest for the distal items, r= .61, p<.01 (proximal items: r=.30, p<.01). The correlation between OTL and performance on the proximal items was r=.28, p>.01 and on the distal items r=.05, p>.01. **No statistically significant effects for OTL on achievement.**
	Comments	General measures of student prior achievement collected at the end of the previous school year were used as covariates in the multilevel analyses (p.31).
Mo, Singh & Chang (2013)	Focus	This study examined the individual, class, and school level variability of the students' science achievement. And it makes a contribution to a better understanding of the OTL variables at classroom and school level in students' science achievement (p.3). This is a correlational study conducted in the USA.
	OTL measure	OTL was measured as a classroom-level factor. Operationally, it included two factors: an indicator of teacher quality (science certification) and an indicator of instructional practice in terms of topic coverage (p. 4). Data from TIMSS 2003 is used.
	Respondents	8th grade students (N=8.544)
	Dependent variable	Science achievement
	Effect size	**The two-class-level OTL variables significantly influenced the class-mean science achievement.** The percentage of variance in student science achievement explained by OTL at the class level (Level 2) was 23,32%.
	Comments	Study includes individual- (students' science- and classroom engagement and students' interests), teacher- (teacher quality and topic coverage), and school-level factors (availability of remedial and enriched courses and the SES of the school (p. 4). Results are not controlled for initial achievement scores.
Niemi, Wang, Steinberg, Baker & Wang (2007)	Focus	This study investigates the instructional sensitivity of a standards-based ninth grade performance assessment that requires students to write an essay about conflict in a literary work. Students were randomly assigned to one of three instructional groups: literary analysis, organization of writing and teacher selected instruction (p. 215).Experimental testing of an assessment's sensitivity to construct-focused instruction is likely to provide stronger validation evidence than OTL data alone (p. 217). This is an experimental study conducted in the USA.
	OTL measure	Sensitivity of a ninth-grade writing performance assessment to different types of standards-based instruction: the differential effects of instruction focused on the organization of writing, literary analysis, or teacher selected goals, controlling for student background variables (p. 218). Sensitivity is measured by data from the district's ninth grade language arts performance assessment made by the students after 8 days of a certain type of instruction.
	Respondents	9th grade students (N=886) & teachers (N=25)
	Dependent variable	Writing performance
	Effect size	The overall performance assessment score shows an advantage of .22 points for the literary analysis group versus the teacher choice group, after controlling for SAT-9 reading scores and language scores, and **this difference is significant**. Scores for students in the writing group were **not significantly different** from scores from students in the teacher choice group (p.226).

	Comments	Results controlled for students' Grade 8 SAT-9 language scores, SAT-9 reading scores, free or reduced-price lunch program status and English language proficiency levels (p. 223, 224).
Oketch, Mutisya, Sagwe, Musyoka & Ngware (2012)	Focus	The primary concern in this paper is to understand some of the classroom-school factors that may explain the persistent differences in achievement between the top and bottom schools. The focus is on time-on-task and curriculum content and whether this explains the difference in performance (p. 19). This is a correlational study conducted in Kenya.
	OTL measure	The effect of active teaching and content coverage on student achievement between low and high performing schools. To conduct this analysis a two-level multilevel model is fitted to evaluate to what degree content coverage, proportion of lesson time spent on active teaching influence student achievement (p.23). Content coverage is measured through the analysis of classroom observation videos. Item response theory was used to calculate test scores, it generated 40 items in each test (p.22).
	Respondents	6th grade students (N=2.437) & teachers (N=72)
	Dependent variable	Mathematics achievement
	Effect size	In the final model shows the effect OTL and time on active teaching on pupil IRT gain score, it controls for pupil, school and teacher characteristics. **Proportion of topic covered (OTL) is positive though not significant.** The proportion of time on active teaching is negative and not significant (p. 29, 31).
	Comments	Model controls for pupil, school and teacher variables. Achievement is tested at two times.
Ottmar, Konold & Berry (2013)	Focus	Examination of the extent to which exposure to content and instructional practice contributes to mathematics achievement in fifth grade. Result suggest that more exposure to content beyond numbers and operations (i.e., geometry, algebra, measurement, and data analysis) contribute to student mathematics achievement, but there is no main effect for increased exposure on developing numbers and operations (p.345). This is a correlational study conducted in the USA.
	OTL measure	Contribution of exposure to specific mathematical content and instructional practice (i.e., geometry, algebra, measurement, data analysis) to mathematics achievement scores. Teachers of sampled children were asked to respond to 24 instructional practice and content items taken from the revised child-level fifth-grade mathematics teacher questionnaire. The fifth-grade mathematics assessment was administered to children using workbooks with open-ended questions (p. 348, 349).
	Respondents	5th grade students (N=5.181), teachers and parents
	Dependent variable	Mathematics achievement
	Effect size	Results indicate that greater exposure to content beyond numbers and operations contributed to higher achievement, **p < .01**. More exposure to numbers and operations or instructional practices **did not significantly contribute** to achievement growth, all p's > .05 (p.351).
	Comments	Results controlled for child and teacher/classroom variables, like students' SES but not for previous student achievement.
Plewis (1998)	Focus	This paper looks at between teacher differences in pupils' mathematics progress from two correlational studies in London schools. We find that the more of the mathematics curriculum covered by teachers, the greater the progress made by pupils in those classrooms (p. 97). Both studies are correlational and conducted in England.
	OTL measure	Effects of curriculum coverage and classroom grouping. The method of measuring curriculum coverage was essentially the same in the two studies. Each teacher completed a checklist for each pupil in the class, the checklist consisting of separate items put into groups such as addition, money, etc., which the teachers ticked if they

		had covered that item during the year with a particular pupil. Thus, we measured coverage of the curriculum experienced by the pupils but reported by their teachers. Each teacher was interviewed about their grouping practices at the end of Year 2 (p. 101) .	
	Respondents	First grade students (N= 776)	Second grade students (N= about 550) & teachers (N= 28)
	Dependent variable	Mathematics achievement	
	Effect size	**The effect size for mean curriculum coverage is 0.11 SD units** accounting for 15% of between teacher variance (**p<0.02**). The effects of classroom grouping on achievement were small. There was some benefit in being in a grouped classroom, with pupils making 0.18 SD units more progress than pupils in the other two types (whole class and individual instruction) of classroom after allowing for the effect of curriculum coverage at the pupil level. The differences in mean curriculum coverage across these three groups were not statistically significant (p. 103, 104).	**The effect size for mean curriculum coverage is 0.18 SD units**, accounting for 65% of the between teacher variance (**p<0.001**). In contrast to study 1, content coverage was lowest for the 'grouped instruction' group. The effect on progress of being in the 'grouped instruction' category was very small and negative. Differences between whole class and individual instruction were not significant (p. 104, 105).
	Comments	Unknown	Results at least controlled for gender and ethnicity proportions in classrooms.
Polikoff & Porter (2014)	Focus	This article is the first to explore the extent to which teachers' instructional alignment is associated with their contribution to student learning and their effectiveness on new composite evaluation measures using data from the Bill and Melinda Gates Foundation's Measurement of Effective Teaching (MET) study (p. 1). This is a correlational study conducted in the USA.	
	OTL measure	The Surveys of Enacted Curriculum (SEC); The surveys define content at the intersection of specific topics and levels of cognitive demand (see Porter, 2002); there are 183 fine-grained topics in mathematics and 133 in ELA. Cognitive demand varies from memorization to application or proof. Application: first decide which topics were taught or not (in a school year); for those taught indicate a) the number of lessons spent on each topic and b) the level of cognitive demand (cell= topic by cognitive demand combination).	
	Respondents	4^{th} and 8^{th} grade teachers (N=701, 327 completed surveys)	
	Dependent variable	Value added measurement in Math and ELA	
	Effect size	When it comes to the zero-order correlations of VAM scores with SEC instructional alignment indices, most of the correlations are not significant. Three correlations with VAM were analyzed: the alignment between instruction and state standards, and the alignment between instruction and state or alternate test. **In those grade, district, subject combinations where the correlations were significant the average was .16 for math and .14 for ELA.**	
	Comments	Most of the zero order correlations were not significant. It should be noted that the independent variable was not the enacted curriculum, but various alignment indicators, e.g. the consistency between SEC and the contents of assessment tests. State and Alternate Assessment VAM (value-added models) scores were used.	

Ramírez (2006)	Focus	This study compared Chile to three countries and one large school system that had similar economic conditions but superior mathematics performance and examined how important characteristics of the Chilean education system could account for poor student achievement in mathematics. One of the results: the Chilean mathematics curriculum covered less content and fewer cognitive skills (p. 102). This study is correlational and is conducted in the USA.
	OTL measure	Content coverage measured by students and teachers' self-reported questionnaires. This study used TIMSS 1998/99 data from Chile, South Korea, Malaysia, the Slovak Republic and Miami Dade Country Public Schools.
	Respondents	8th grade students (N between 1.356 and 6.114 per country) and their Mathematics teachers.
	Dependent variable	Mathematics achievement
	Effect size	In Chile, 73% of the students were taught by teachers who emphasized basic mathematics content. In the comparison jurisdiction, this proportion was substantially smaller (6%, 12%, 19% and 33%). In Chile, **content coverage was significantly related to mathematics performance**. This relationship held true after controlling statistically for schools' socio-economic index and type of administration, 4.8, p <.05).
	Comments	Results are controlled for schools' socio-economic index and type of administration (public/private), but not for prior achievement.
Reeves (2005)	Focus	This thesis investigates whether the existing South African policy approach is supported through research, or whether, in accordance with the international evidence, 'Opportunity-to-Learn' (curriculum content and skills actually made available to learners in classrooms) has a greater effect on achievement (than 'type of pedagogy') and is therefore a policy variable worth taking more seriously for narrowing the gap in achievement between South African learners on different socio-economic backgrounds (p. iii). This is a correlational research conducted in South Africa.
	OTL measure	Four OTL dimensions: content coverage by cognitive demand, content exposure, curricular coherence and curricular pacing, measured by lesson observations, teacher survey interviews, teachers' year or term plans and students questionnaires (partly items from TIMSS) and students' workbooks and reports.
	Respondents	6th grade students (N=1.001) and their mathematics teachers.
	Dependent variable	Mathematics achievement
	Effect size	The study's findings do not confirm the assumption that in relation to achievement gain, OTL is more important than 'type of pedagogy'. The results show that OTL and pedagogy variables both significantly affect achievement (p. 230). The variable that had the highest correlation with achievement gain was the level of cognitive demand (a correlation co-efficient of 0.28).
	Comments	Results controlled for individual learner background variables. This study uses achievement gain scores.
Reeves (2012)	Focus	Research has shown that rural high school students in the United States have lower academic achievement than their nonrural counterparts. The evidence for why this inequality exists is unclear, however. The present study takes up this issue with a narrowing of the focus. Using the database of the Educational Longitudinal Study of 2002-2004, the author investigates reasons for the rural achievement gap in mathematics during the last 2 years of high school. His approach focuses on the geographic disparities in the opportunity to learn advanced math (p. 887). This is a correlational study conducted in the USA.
	OTL measure	The supply-side factors of OTL, such as school offerings of advanced math units, restriction on student admission to advanced math courses, or the quality of advanced

		math instruction. This will be measured using survey regression models focused on comparative effects of family SES on course taking in different geographic locations and separate regression models of math achievement gain will be estimated for each type of school location.
	Respondents	10^{th} grade (and two years later, 12^{th} grade) students (N=11.170)
	Dependent variable	Mathematics achievement
	Effect size	In Model 3, we find that the addition of the opportunity-to-learn variable – total advanced math units taken – not only has a large effect on the math achievement gain, but it also accounts for more than two third of the residual rural gap and reduces the remaining gap to nonsignificance (p. 901).
	Comments	Results controlled for 10^{th} grade achievement, student demographics, private school attendance, school size, family SES, and friends' educational engagement and aspirations (p.899).
Reeves, Carnoy & Addy (2013)	Focus	This paper estimates the effect of OTL on students' academic performance using rich data we gathered on the teaching process in a large number of South African and Botswana Grade 6 classrooms (p. 426). This is a correlational study conducted in Africa.
	OTL measure	Curriculum coverage, including: content coverage, content exposure and content emphasis). The data comes from student notebooks and videotaped mathematics lessons.
	Respondents	6^{th} grade students (N > 5.000) and teachers (N= 116)
	Dependent variable	Mathematics achievement
	Effect size	The study's estimates suggest that in many of the South African classrooms the relation of additional lessons on test items to test score gains, although positive, is **not statistically significant**. The test score gain on items in Botswana classrooms is generally **negatively related** to the number of lessons given by teachers on each test item (p. 432).
	Comments	Results are controlled for pre-test scores, but not for students' SES.
Schmidt (2009)	Focus	Exploration of the relationship of tracking in eighth grade to what mathematics topics are studied during eighth grade (content exposure) and to what is learned during the year as well as to what is achieved by the end of eighth grade (p.6). This is an experimental study conducted in the USA.
	OTL measure	Effect of tracking in different types of courses: regular, pre-algabra and algebra. Each of the sampled school defines a track in the sense of providing different content opportunities to learn mathematics. TIMSS surveyed the mathematics teachers of the sampled classes.
	Respondents	7^{th} grade students (N= 3.886), 8^{th} grade students (N=7.087), 7^{th} grade teachers (N= 127) and 8^{th} grade teachers (N= 241)
	Dependent variable	Mathematics achievement
	Effect size	In tracked schools, the algebra track was statistically significantly different from the other two tracks (p<.0001), it covered content slightly over one grade level higher (1.09) than the regular track and almost one (.92) than the pre-algebra track (p.16). For algebra classes the 70-point difference in mean achievement between those in tracked schools versus non-tracked schools is significant (**p<.003**), but the differences in mean achievement for the other two types of courses are not significant. Across the non-tracked schools there were no significant differences in eighth grade achievement for the three different type of courses (p<.38) (p.21).
	Comments	The track designation was included as a dummy variable at the classroom level. The model also included several covariates at each of the levels in the design. The student-level included racial identity and SES, the class-level included 7^{th} grade pre-measure,

		mean SES and track, the school-level included the school-level mean SES, percent minority enrolment, location and size of the school (p.23).
Schmidt, Cogan, Houang & McKnight (2009)	Focus	Analyses that explores the relationship between classroom coverage of specific mathematics content and student achievement as measured by the TIMSS-R international mathematics scaled score (p.i). This is a correlational study conducted in the USA.
	OTL measure	Variation in content coverage across a set of districts and states and relating it to cross-district/state variation in achievement. The study uses IGP, ''international grade placement'', that provides an indication of the conceptual complexity for each topic. The data came from the teacher questionnaire in which they indicated the number of periods of coverage associated with each of a set of topics.
	Respondents	8th grade students (N=36.654) and their mathematics teachers
	Dependent variable	Mathematics achievement
	Effect size	Districts that had a higher average value on the IGP index also had a correspondingly higher mean achievement (R2= 67 percent, **p<.01**)). To test the effect of OTL on achievement controlling for SES, both variables were included in the same district level regression model. Both were related to achievement (R2 = 82 percent, **p< .0002**).
	Comments	Results controlled for SES at all three levels and prior achievement.
Snow-Renner (2001)	Focus	Academic achievement and opportunity to learn were studied using data from the 195 TIMSS for Colorado students at the elementary level. The study used a comprehensive definition of OTL that includes content coverage, curricular focus, duration of instruction, and instructional strategies. The implications for using large-scale measures to indicate how fairly educational opportunities were distributed were studied in a context of comparative accountability measures (p.1). This is a correlational study conducted in the USA.
	OTL measure	Content coverage measured in terms of curricular focus (number of topics taught b teachers) an topic coverage. The measures from student achievement scores and teacher surveys are mapped specifically onto six different subtopic: whole number; fractions and proportionality; measurement, estimation, and number sense; data representation, analysis and probability; geometry; and pattern, relations, and functions. Due to lack of a reasonable level of consistency reliability, the geometry and patterns subscales were omitted from the remainder of the study (p. 8, 9).
	Respondents	third and 4th grade students (N=2.163) and their teachers
	Dependent variable	Mathematics achievement
	Effect size	The only variable that correlates significantly and positively across grade levels with all four achievement subscales is the curricular focus variable **(8/8)**. For the other variables concerning topic coverage, correlations are inconsistent by grade level. **No fourth grade classes showed any significant relationships between achievement and topic coverage. The third grade classes showed significant correlations for 6 of the 12 variables, all positive (overall 6/24 significant).** In contrast, fourth grade achievement correlated most highly and significantly with variables measuring instructional practices rather than topic coverage (p.13, 14).
	Comments	Results not controlled for other independent variables or co variables.
Tarr, Ross, McNaught, Chávez, Grouws, Reys, Sears & Taylan	Focus	American curricula seems more skills oriented, more repetitive and less conceptually deep than those of nations that score better than America on TIMSS. This research-study focuses on the question whether there are differences in mathematical learning when students study from an integrated approach textbook and when they study from an subject-specific textbook. And what are the relationships among curriculum type, fidelity of implementation and student learning (p. 1, 2). This is a correlational study

(2010)		conducted in the USA.
	OTL measure	Influence of curriculum type: integrated approach textbook vs. subject-specific textbook. The study uses classroom visits, teacher surveys, textbook diaries, project developed tests and standardized tests. Another factor is what they called OTL, including the percentage of textbook lessons taught by the teacher during the year, the Extent of Textbook Implementation index, the seating arrangement of observed lessons and the dominant level of student engagement in observed lessons. Lastly, the factor implementation fidelity, including Textbook Content Taught index, Content fidelity rating and the Extent of Textbook Implementation index (p. 19).
	Respondents	8th grade students (N=2.621) and teachers (N=43)
	Dependent variable	Mathematics achievement
	Effect size	Curriculum type is positively related in the three different test when no other variable is disregarded, but only 2 of the 3 are significant (**r=.304, p<.05; r=.518, p<.001; r=.264, p>.05**). OTL, is in all three tests positively and significantly related to student outcomes (**r=.388, p<.01; r=.370, p<.05; r=.291, p<.05**). Fidelity, is in non of the test significantly related to student outcomes (**r=-.189, p>.05; r=-.085, p>.05; r=-.115, p>.05**) (p. 23).
	Comments	Results controlled for prior achievement. Correlations between student outcomes and ten other variables are measured partialling out the other variables one at a time.
Törnroos (2005)	Focus	Relation between OTL and mathematics achievement in which OTL is approached in three ways. Firstly, it was measured as the proportion of textbooks dedicated to different topics. The second approach was based on the data given by teachers in TIMSS 1999. The third approach involved an item-based analysis of the textbooks (p. 320). This is a correlational study conducted in Finland.
	OTL measure	Content coverage divided into three variables: Proportion of textbooks dedicated to topics (INBOOKx), what has been taught by teachers (TAUGHTx) and the proportional analysis of the textbook content (CONTENTx).
	Respondents	7th grade students, teachers and textbooks (N=9)
	Dependent variable	Mathematic achievement
	Effect size	For textbook K only the variable TAUGHT had a statistically significant correlation with achievement **(1/3)**. Textbook P showed no statistically significant correlations between OTL and achievement **(0/3)**. For textbook MM the variable INBOOK had what were clearly the highest correlations with achievement **(1/3)** (p. 321).
	Comments	This analysis was based on students' actual achievement instead of achievement gains over a specific time period (p.321). Results are not controlled for other variables. However, a distinction is being made between raw and standardized scores.
Wang (1998)	Focus	This study investigated the relationship between students' OTL and their science achievement. Hierarchical linear modelling was used to analyze OTL variables at two levels of instructional processes: the classroom level and the student level (p. 137). This is a correlational study conducted in the USA.
	OTL measure	Eight OTL variables covered by four constructs: content coverage, content exposure, content emphasis, and quality of instructional delivery. The latter is rather broad including also for example teacher preparation and equipment use. Science achievement is measured in both a written test and a hands-on test. In addition, teachers were interviewed about content coverage, activities and their prediction of how well their students would do on post-test. The teachers also provided copies of all the material they used as well as student daily attendance lists (p. 141).

	Respondents	8th grade students (N=623) and science teachers (N=6)
	Dependent variable	Science achievement
	Effect size	It was found that OTL variables were significant predictors of both written and hands-on test scores even after students' general ability level, ethnicity, and gender were controlled. Content exposure was the most significant predictor of students' written test scores, and quality of instructional delivery was the most significant predictor of the hands-on test scores (p. 137). Written tests: Content Exposure (β=11.1, SE=5.4), Content Coverage (β=10.6, SE=9.9) and Quality of Instructional Delivery (β=5.8, SE=4.0). Hands-on tests: Content Exposure (β=14.2, SE=7.0), Content Coverage (β=25.4, SE=12.8) and Quality of Instructional Delivery (β=10.8, SE=4.9) (p.149).
	Comments	Results are controlled for students' general ability level, ethnicity, and gender, but not for students' SES. Content emphasis was omitted from the analyses because of its high correlation coefficients with content coverage, content exposure, and quality of instructional delivery (p. 152).
Wang (2009)	Focus	This study empirically examined a subset of children from low-income families to determine whether African American and Caucasian students have differential opportunity to learn mathematics and the extent to which opportunities to learn predict gains in mathematics achievement at kindergarten (p. 295). This is a correlational study conducted in the USA.
	OTL measure	OTL variables representing maths instructional time, maths instructional method (three variables), and maths instructional emphasis (two variables). Students were assessed in maths skills and knowledge both kindergarten entry and exit, and teachers were asked to complete a survey that included 48 items relating to maths OTL (p.297) .
	Respondents	Kindergarten students who lived below the poverty line (N=1.721)
	Dependent variable	Mathematics achievement
	Effect size	OTL was found to predict maths achievement of African American and Caucasian kindergartners from low-income families. Both groups showed only 1 statistically positive significant correlation with achievement, which is the OTL variable 'Emphasis: Telling time, estimating quantities and coin values accurately 1-2 times per week'. For African American children the variable 'Method: Used math manipulatives at least 1-2 times per week' is negatively but statistically significantly related to math achievement (b= -1.23) **(Total: 2/6 significant correlations)**. For the Caucasian students there a no other significant correlations **(Total: 1/6 significant correlations)**.
	Comments	Results controlled for mathematics achievement at kindergarten entry, student age, student gender, and full-day vs. half-day kindergarten programs.
Wonder-McDowell, Reutzel & Smith (2011)	Focus	The purpose of this study was to explore the effects of aligning classroom core reading instruction with the supplementary reading instruction provided to 133 struggling grade 2 readers. A 2-group, pre-posttest true experimental design was employed in this study conducted in the USA (p. 259).
	OTL measure	Influence of aligned and unaligned supplementary reading instruction after a maximum of 20 weeks. Effect is measured by pre- and posttest with a focus on reading fluency, word identification, word attack and reading comprehension.
	Respondents	Second grade students (N=133) and teachers (N=12)
	Dependent variable	Reading achievement
	Effect size	Struggling readers in both the aligned and unaligned supplementary reading instruction groups made significant growth across all measures from pretest to posttest during the treatment period. The eta-squared effect size indicated **for all four variables a small but**

		statistically significant positive effect of aligning supplementary reading instruction on student's growth. The effect size for reading fluency is .17 (p<.001), for word identification .08 (p<011), for word attack .13 (p<.001) and for reading comprehension the effect size is .18 (p<.001) (p.272).
	Comments	Demographic variables of gender, reading achievement, ethnicity, English learner status, and free and reduced-price meals qualification are taken into account, there were no significant differences between both groups.

Appendix 3: Changes compared to the original study

Deleted from the original study because of incomparability

- The six meta-analyses do not form part of this study. Primary empirical studies are, in contrast to meta-analyses, comparable when it comes to certain study characteristics like subject matter and age group. A meta-analysis is transcending the study characteristics that will be highlighted and therefore not useful for this study.

Deleted from the original study because the article falls not within the predefined inclusion criterion about the year of publication

- Gamoran, A. (1987). Instruction and the effects of schooling. Paper presented at the annual meetings of the American Sociological Association.
- Winfield, L.F. (1987). Teachers' estimates of test content covered in class and first-grade students' reading achievement. *The Elementary School Journal, 87(4), 436-454.*
- Yoon, B., Burnstein, L., Chen, Z. & Kim, K.-S. (1990). *Patterns in teacher reports of topic coverage and their effects on math achievement: comparisons across years* (CSE Tech. Rep. No. 309). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Yoon, B, Burnstein, L. & Gold, K. (1991). *Assessing the content validity of teachers' reports of content coverage and its relationship to student achievement* (CSE Rep. No. 328). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

The four articles, that are deleted because they are outdated, do meet the other inclusion criteria. Nevertheless, they are not useful for this study due to the fact that no complete literature search is done for articles published before 1995. These articles were found by chance. Statistical comparisons concerning the year of publication are therefore not entirely reliable, because there are probably more articles about OTL published in those years. Instead of doing a complete literature search for articles about OTL from before 1995, it has been decided to delete these articles. This study is meant for deepening rather than widening.

Deleted from the original study because the OTL variable was not content coverage

- Cai, J., Wang, N., Moyer, J. C., Wang, C., & Nie, B. (2011). Longitudinal investigation of the curricular effect: An analysis of student learning outcomes from the LieCal Project in the United States. *International Journal of Educational Research, 50* (2), 117-136.
- Calhoon, M. B., & Petscher, Y. (2013). Individual and group sensitivity to remedial reading program design: Examining reading gains across three middle school reading projects. *Reading and Writing, 26*(4). 565-592.
- Grouws, D. A., Tarr, J. E., Chávez, O., Sears, R., Soria, V. M., & Tatlan, R. D. (2013). Curriculum and implementation effects on high school students' mathematics learning from curricula representing subject-specific and integrated content organization. *Journal for Research in Mathematics Education, 44*(2), 416-463.

- Roncagliolo, R. (2013). *Time to learn mathematics in public and private schools: Understanding difference in aspects of the implemented curriculum in the Dominican Republic* (Dissertation). Retrieved from ProQuest Dissertations and Theses.
- Tarr, J. E., Reys, R. E., Reys, B. J. Chávez, Ó., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grade mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education, 39*(3), 247-280.

These five articles do meet all the other inclusion criteria. However, this study focuses only on content coverage as OTL variable. Studies that focus on more than one variable, including content coverage, are included in this research. Studies that investigate only other OTL variables than content coverage are excluded from this study. In the original study content coverage was already the leading variable, but the exclusion of studies was less strict.

Adapted in this study

- Plewis, I. (1998). Curriculum coverage and classroom grouping as explanations of between teacher differences in pupils' mathematics progress. *Educational Research and Evaluation, 4*(2), 97-107.

This article describes two studies concerning OTL. In this analysis, both studies are considered as two separate units.

Appendix 4: Critical values of the χ^2 distribution

Table 14: Critical values of the χ^2 distribution

p df	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188