

Chapter 2

Conceptualization

Jaap Scheerens

Abstract In the second chapter a conceptual analysis of Opportunity to Learn (OTL) is given, covering also related terms, such as instructional alignment and test preparation. The OTL issue is highlighted from three educational research traditions: educational effectiveness research, curriculum research and achievement test development. The conceptual analysis leads to pinpointing OTL as a specific type of alignment in educational systems; a taxonomy of alignment forms is presented. Next, different facets of the way OTL is measured empirically were discussed. The conceptual analysis is given further theoretical depth, by discussing De Groot's (Begrip van evalueren. Vuga, 's-Gravenhage, 1986) integrative model of didactic and evaluative operationalization. Reflecting on this model brought the alignment issue in a systemic perspective, leading up to the conjecture that alignment, OTL and test preparation aim for integration in organizational structures that are often characterized as loosely coupled, which might explain sub-optimal effects of these policies.

Introduction

At first glance “opportunity to learn” would seem to be one of the few concepts in educational science that you could clarify to you mother or grandmother in two minutes. It addresses the expectation that students will do better on educational content tested when that content has actually been taught, which almost sounds like a truism. Throughout this report we will remain close to this basic clarification, since we are not here to complicate matters unnecessarily. As this study is a

J. Scheerens (✉)
University of Twente, Enschede, The Netherlands
e-mail: j.scheerens@utwente.nl

J. Scheerens
Oberon, Utrecht, The Netherlands

© The Author(s) 2017
J. Scheerens (ed.), *Opportunity to Learn, Curriculum Alignment and Test Preparation*, SpringerBriefs in Education,
DOI 10.1007/978-3-319-43110-9_2

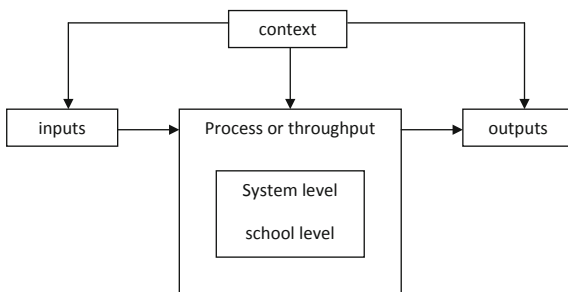
research review we shall also encounter the basic and simple conception of opportunity to learn (abbreviated as OTL) in the empirical studies that will be analyzed. The correlation between a measure of OTL and cognitive achievement in school subjects, like mathematics, science and mother tongue language, is the key parameter of investigation. Nevertheless the exploration of the literature shows complexity that goes beyond the basic definition. The OTL issue can be situated in at least three different traditions of educational research and development, with corresponding differences in research orientation, shows considerable variability in the way it has been defined and operationalized, and has different contexts of practical application as well (e.g. national educational policy and school level teaching). In this chapter a “conceptual map” of OTL will be presented.

Building Blocks for a Conceptual Framework

OTL is a construct that depends on the *alignment* of educational goals or standards, actual teaching and assessment (measurement of student achievement). These elements can be positioned in a basic system model of education, which is often used to define educational effectiveness.

The elementary design of educational effectiveness research is the association of hypothetical effectiveness enhancing conditions (OTL being one of these) and output measures, mostly student achievement. The basic model from systems theory, shown in Fig. 2.1, is helpful in clarifying this design. The major task of educational effectiveness research is to reveal the impact of relevant input characteristics on output and to “break open” the black box in order to show which process or throughput factors “work”, next to the impact of contextual conditions. The model, shown in Fig. 2.1, can be used at different levels of aggregation. In the figure this is indicated by mentioning three levels in the central box of the model: the level of a national educational system, the school level and the level of the instructional setting, often indicated as the classroom level. The three levels are nested, in the sense that schools function within an educational system at national level and classrooms function within schools.

Fig. 2.1 A basic systems model on the functioning of education



In terms of this model the alignment between standards, actual teaching and student achievement can be seen as the association of inputs (e.g. national standards), processes (teaching) and outputs (student achievement). Accordingly, OTL can be characterized as the alignment between inputs and teaching processes, as the alignment between teaching processes and student achievement, or as the alignment of standards and output measures, mediated by teaching processes. An example of a relevant context variable is the degree of centralization of an educational system. To the degree that the educational system is centralized, national standards, or a national curriculum, are likely to be more detailed and prescriptive in the way they are supposed to be applied by schools. When an educational system is more decentralized national curricula might be just rudimentary, consist of quite general goals, or even be totally absent.

The educational effectiveness perspective is just one of three research and development orientations in which OTL is approached in a specific manner. The other two perspectives are the logic of curriculum research and test preparation. In this study our emphasis will be on the educational effectiveness perspective; so this context for OTL research will be explained first.

OTL in the Context of Educational Effectiveness Research

In educational effectiveness research OTL is one of a series of malleable, effectiveness enhancing conditions at national system, school and classroom level that are expected to be positively associated with student achievement, also when outcomes are adjusted for student characteristics like previous achievement, scholastic aptitude and socio economic status. Other malleable variables addressed in educational effectiveness research are indicated in the overview presented in Table 2.1, cited from Scheerens (2014).

Table 2.1 Summary of effectiveness enhancing teaching variables by Muys et al. (2014), adapted from Scheerens (2014)

Teaching effectiveness, Muys et al. (2014)
Opportunity to learn
Time
Classroom management
Structuring and scaffolding, including feedback
Productive classroom climate
Clarity of presentation
Enhancing self-regulated learning
Teaching meta-cognitive strategies
Teaching modelling
More sophisticated diagnosis
Importance of prior knowledge

In educational effectiveness research OTL has been used as an independent variable defined mostly at school and classroom level. With the development of international assessment studies country level definitions of OTL have also been used. A defining characteristic of OTL studied from an educational effectiveness perspective is that measured student achievement is the dependent variable.

OTL in Curriculum Research

Curriculum research, rather than educational effectiveness research, forms the intellectual heritage of OTL. OTL in curriculum research shares the systemic perspective with the more recent multi-level studies in educational effectiveness (Scheerens 2016). The research orientation in curriculum research is broader than in educational effectiveness research. In the curriculum context alignment is addressed in its broadest sense, including sometimes “alignment” with measured student achievement, interpreted as the “realized curriculum”, but not limited to that. The building blocks for our conceptual framework on OTL that were previously mentioned (standards, actual teaching and student outcomes) have specific terminology in curriculum research, where one speaks of the intended, implemented and realized curriculum. In curriculum research alignment between national curriculum standards (intended curriculum), intermediary elements, such as school level standards and textbook content, and taught content is studied in its own right, without necessarily involving student outcomes.

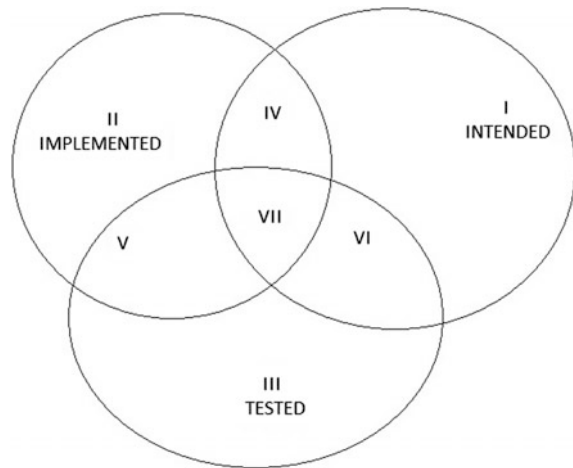
Alignment between standards, intermediary elements, teaching and assessment of student outcomes has different interpretations when considering the association between pairs of elements.

- (1) *Alignment between curriculum goals or standards and intermediary elements* such as school standards or textbooks could be considered in terms of construct validity; the key question is whether textbooks or school standards provide a “covering” representation of the national standards. Assuming that national goals or standards are likely to be defined in more general terms than are the content elements in school standards and textbooks, the analogy with construct validity seems more appropriate than content validity, which would presuppose matching of elements from two sources of comparable specificity. In the case of construct validity expert panels would be needed to decide whether school standards or textbook content could be seen as adequate representations of the more general goals, or national standards.
- (2) *Alignment between curriculum goals and standards and teaching* (i.e. *the implemented curriculum*). Here the same reasoning could be applied as in case 1, in the above. The actual feasibility of assessing this kind of alignment would strongly depend on the national standards being sufficiently specific; in addition empirical methods to observe or otherwise measure teaching behavior in practice would be required.

- (3) *Alignment between intermediary elements (school standards and textbooks)* would allow for a more straightforward consistency check, based on content analysis of the school standards and textbooks and matching with measures of content covered by teachers. Here the practical reason for carrying out a consistency check could be the choice of the most suitable textbook, given school standards
- (4) *Alignment between intermediary elements and assessment instruments.* Here content elements of the intermediary elements would be matched with the content elements that make up the assessment instrument. This might be done either based on actual test items or on (more general) content elements derived from the analytic frameworks on which the test is based. Such frameworks usually consist of two dimensional matrices, specifying cognitive operations required in relationship to content elements. The context of application might be test validation or analyzing the opportunities to learn that are stimulated by school standards or textbooks.
- (5) *Alignment between teaching (implemented curriculum) and assessment content.* This would have essentially the same interpretation and contexts of application as described with point 4.
- (6) *Alignment between general goals or national standards and assessment content.* As with the alignment discussed in point 2, the feasibility of this approach would strongly depend on the specificity of the national standards.
- (7) *“Alignment” of any other of the main elements to student achievement outcomes (the realized curriculum).* This kind of alignment refers to the most common definition of OTL. The term alignment is questionable in this case, because the association, although mostly just measured by means of correlational measures is likely to be interpreted as causal. The most frequent application is the one between content covered in teaching and achievement results. Contexts of application are: establishing the predictive validity of OTL measures and assessing school or teaching effectiveness. In the latter sense curriculum research and educational effectiveness research overlap.
- (8) More complex models of alignment, where intermediary elements may be studied as mediators of higher level elements (examples will be provided in Chap. 3).

In this taxonomy of alignment types, when national standards, intermediary elements (school standards and textbooks) and assessment instruments and measures are the basic elements, the emphasis has been on matching and consistency. Pelgrum (1989), presents a conceptual framework in which mismatches and deficiencies, next to matches, are given explicit attention. His work took place in the context of international comparative assessment studies by the IEA (International Association for the Evaluation of Educational Achievement), in which variability between countries in the way the international assessment test corresponded to national intended and implemented curricula, was scrutinized from the perspective of “fair” comparison.

Fig. 2.2 Venn diagram of intended, implemented and tested curriculum, from Pelgrum (1989)



The presentation of Pelgrum’s model is cited in Fig. 2.2 (Pelgrum 1989, p. 17). The numbered areas in Fig. 2.3 are described as follows:

“I + IV + VI + VII: what students should learn.

II + IV + V + VII: what is actually taught at school.

III + V + VI + VII: what is tested.

I: what students should learn, but is actually not taught at school, and not tested.

II: what actually is taught at school, but not tested and not part or what students are supposed to learn.

III: what is tested, actually not taught at schools, and not part of what students are supposed to learn.

IV: what students should learn and what is actually taught at school, but not tested.

V: what actually is taught at school and tested, but is not part of what should be learned.

VI: what students should learn and is tested, but is actually not taught at schools.

VII: what students should learn, what is tested and taught.” (ibid., 17).

The theoretical principle behind these analyses of consistency between the various facets of curriculum can be indicated with the term “coupling”. Analyses that tend to underline deficiencies could be seen as manifestations of “loose coupling” in educational organizations (Weick 1976); the positive alternative of good integration between the curriculum facets can be indicated with the term “alignment”. Successful OTL is an example of alignment, fallible OTL can be seen as a manifestation of loose coupling.

OTL as Test Preparation

In this section we shall start out with an orientation on the process of educational achievement test development. As we shall see test development follows the same kind of specification process, from general goal descriptions to test items, as were encountered in curriculum analysis and development. When comparing curriculum development and test construction we can establish, first of all that they have the first and last step of the development process in common, the first being a national curriculum with general goals and national standards, and the last step being the assessment instruments. Most interesting are intermediary “products”. In curriculum research we encountered school standards, textbooks and implemented curriculum as intermediary steps. Analyzing test construction shows other kinds of intermediary products. When discussing test preparation as an interpretation of OTL these intermediary products are quite interesting. Holcome’s: “taxonomy of score inflation”, illustrates what is meant by intermediary products in test development (Holcombe, 2011).

The term “score inflation” refers to the context of application of this taxonomy, which is teaching to the test. Each of the decisions in test design (like specifying subsets of standards and material to be covered within standards) is seen as narrowing the domain for testing and creating opportunities for teaching to the test. The subsequent decisions in test development concern content specification but also choice of representations, such as item formats.

The specification process in test development for a particular subject could be seen solely from a content perspective. The deductive steps then go from major domains of a discipline, to subdomains, to more specific topics and ultimately to item content. However, at the more detailed levels, the level of topics and test items, a second dimension is usually added, in the form of the cognitive demand of the topic or item. Topics are thus defined as a combination of the specification of a content element and a particular psychological operation. The cognitive demand dimension can be arranged from simple to more complex cognitive operations. See Porter et al. (2011, 104).

In the test frameworks for PISA the cognitive operation dimension is further sub-divided in terms of process categories and cognitive demand. Next, a context dimension and a response type dimension are distinguished to further characterize test domains and test items. The context dimension consists of a personal, societal, occupational and scientific sub dimension.

So what does it mean that intermediary specification levels in curriculum and test design have quite similar analytic structures consisting of specification of content and psychological operations with a certain demand or difficulty level? Obviously this facilitates empirical research on different types of curriculum alignment, see for example Porter et al. (2011). Perhaps more interesting is to further reflect on implications for OTL optimization. Here the attention would go particularly to the association between teaching content and test content. The question is whether “test preparation” can be seen as a constructive and “legitimate” way to optimize OTL.

Traditionally this kind of alignment has the unfavorable connotation of “teaching to the test”. But, perhaps, when certain technical requirements of tests are met, specific ways to direct teaching to these tests are not so bad. We shall return to these questions after having further analyzed the communalities and differences between OTL from the curriculum perspective and test preparation facilitated by test characteristics. In the next section an integrative model of “didactic and evaluative specification” (De Groot 1986) will be discussed to try and make further progress on these issues.

An Overarching Model of “Didactic and Evaluative Specification of Educational Goals”

De Groot (1986) describes the development of curriculum programs as the result of a process in which policy goals, background characteristics of students and societal demands are the key inputs to choose general goals, and create an overall vision of how to attain these goals. In a subsequent step of specification, goals are formulated as attainable end-terms (effects); “standards” in more contemporary terminology.

In Fig. 2.3, these steps are represented with A, B and C, in the upper part of the figure. Next the specification process splits up in two directions: “didactic operationalization” (D) and “evaluative operationalization (E)”. The didactic operationalization leads to a concrete plan in the form of school standards and teaching methods, which in a next step is brought into practice (the implemented curriculum). The evaluative specification leads to the design of test instruments and ultimately to test items, norms, and decision rules about success or failure. All relationships in the figure, A through H, are indicated as “coverage problems”; the total of specifications at a lower level should cover the main themes of a higher level. Because higher level descriptions are in broader terms De Groot prefers the analogy of construct validation to judge the success of coverage of goals by curriculum elements and test frameworks at lower levels to content validity. Content validity would be theoretically adequate if the higher level goal formulation would be a precise collection of elements, and a test a representative sample from those elements. However, according to De Groot, educational goals at higher level are more than collections of content elements, because they may also refer to general skills, like problem solving or social skills. And this means that, ultimately, expert judgment is required to assess the content validity of lower level elements, like textbooks, test frameworks and tests. Relationship H in Fig. 2.3 is crucial, it refers to our basic definition of OTL: the degree to which the content tested has actually been taught.

De Groot’s framework underlines the analogy between curriculum and test design, and offers criteria to determine the quality and alignment of these two construction processes. In the recognition of vertical coverage in the didactic and the evaluative column, and “horizontal consistency” between the two columns in

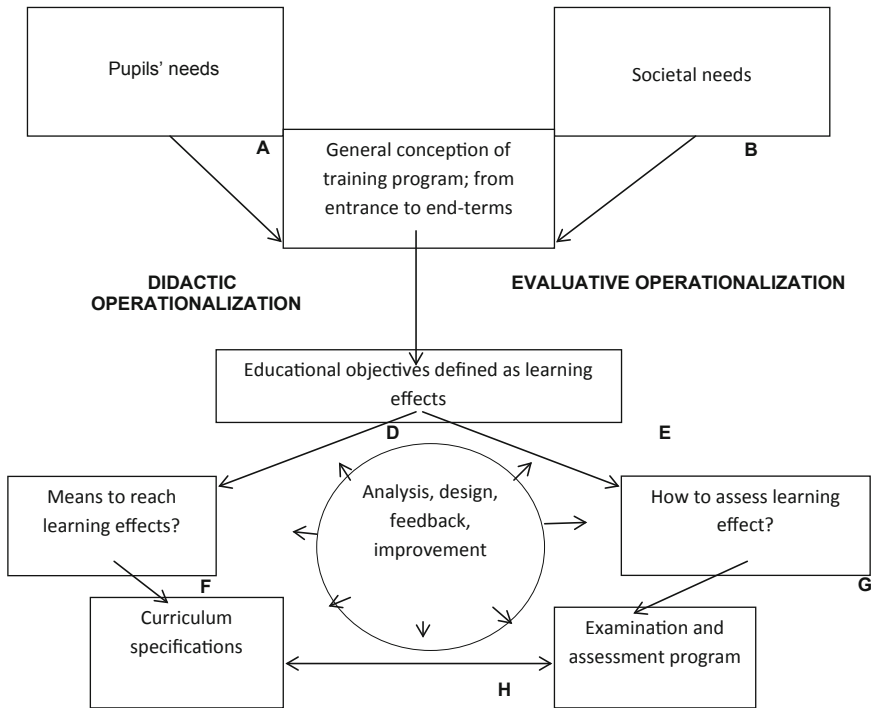


Fig. 2.3 Two kinds of operationalizations of educational goals, adapted from De Groot (1986)

the form of OTL. A few more practical issues inspired by this integrative theoretical framework concern the way the two processes should be organized over time (which should be done first?), whether didactic and evaluative operationalization should be carried out by different, independent organizations and finally, whether alignment would not be served by streamlining the organizational conditions.

Should Evaluative Specification Precede Didactic Specification or Vice Versa?

De Groot argues that evaluative operationalization should happen first, because curriculum design needs verifiable learning effects to adequately resolve issues of instrumentality, in other words constructing means that are adequate to reach goals and intended effects. If the evaluative specification would follow the didactic specification there would be too big a chance of pressure to adapt tests to preferred methods (goal displacement).

Should Evaluative Specification and Didactic Specification be Carried Out in Different Organizations?

According to De Groot evaluative and didactic specification should be independent. His main argument is that curriculum developers lack the know-how to carry out test construction. Next, in high stakes evaluative applications, like examinations, independence is required to guarantee objectivity. In actual practice various organizational units are often involved in specific phases of curriculum development and test development. Developers of teaching methods and textbooks are often independent firms operating outside the public sector; the same may apply to test developers.

From De Groot's analytic model, but also from our earlier presentation of alignment issues, it seems that what we have are two operationalization processes that are quite similar. From a theoretical perspective, but also from the point of practical application it is therefore challenging to think further about a more efficient approach in organizational structures that might be more aligned and less "loosely coupled".

How Feasible is Optimizing Alignment in a Leaner Organizational Structure?

To a degree, alignment in educational systems, as discussed so far, is a remedy to a self-created problem. Particularly in the curriculum development column in Fig. 2.3 organizational units at various levels are involved in the process of what De Groot describes as "didactic operationalization". At national level priorities, general goals and overall time tables are established by either the central administration or national institutes that operate closely to the central administration. Depending on the degree of centralization of the system and the specificity of the national curriculum, at intermediary level various organizational units may have a prominent role in the process of didactic operationalization as well: textbook writers, educational support organizations, school districts and school boards. These intermediary units develop "intermediary curricular elements", like district standards, textbook content coverage, and school standards, creating as many areas where alignment becomes an issue. Again, depending on the specificity of the intermediary elements and the autonomy of teachers, the teachers will have more or less space to make their own choices in what is actually taught. So, in summary, there is a lot of need for vertical coordination in the didactic specification column. Looking once more at the question of horizontal alignment, i.e. the correspondence between elements in the didactic specification column and elements in the evaluative specification column, we saw that De Groot argues for a leading role of test development. Evaluative specification should precede curriculum specification because concrete and specific end terms (i.e. ultimately collections of test items) are needed to guide

the curriculum development process. It is rather questionable whether such kind of interplay and coordination between test development and curriculum development is frequently realized in practice. If it is not realized another alignment issue arises, creating, most probably important discrepancies between what is taught and tested; in other words limited OTL. It is important to realize that the quest for alignment in educational systems, of which OTL is a specific issue, happens in a context where structural arrangements tend to be loosely coupled.

The question is how matters could be improved, first of all “in theory” and secondly in practice, when all kinds of contextual conditions of a structural and cultural nature should be taken into consideration.

Hypothetical Solutions to the Alignment Problem in Educational Systems

Two ideal type scenarios will be addressed in the next section: centralism and synoptic planning, and retroactive planning, combined with high stakes accountability.

Centralism and Synoptic Planning

Although, during the last two decades, there has been a strong tendency in many countries to decentralize educational systems and provide more autonomy to lower levels (schools in particular), some previously decentralized countries like the UK and the USA have gone in the other direction. In the UK national programs for numeracy and literacy were developed and implemented, and in the USA Common Core Standards have been developed. Explicit national curriculum standards provide clear direction for both didactic and evaluative operationalization. At the very least it opens the possibility to empirically verify the alignment between, for example, the national standards and the contents of assessment instruments.

Rational, synoptic planning can be seen as the theoretical background of national curriculum planning.

The ideal of “synoptic” planning is to conceptualise a broad spectrum of long term goals and possible means to attain these goals. Scientific knowledge about instrumental relationships is thought to play an important role in the selection of alternative ways to realize these goals.

The main characteristics of synoptic planning as a prescriptive principle conducive to effective (in the sense of productive) organizational functioning, as applied to education, are:

- “proactive” statement of goals, careful deduction of concrete goals, operational objectives and assessment instruments;

- decomposition of subject-matter, creating sequences in a way that intermediate and ultimate objectives are approached systematically;
- alignment of teaching methods (design of didactical situations) to subject-matter segments;
- monitoring of the learning progress of students, preferably by means of objective tests.

According to this model curriculum development is seen as a deductive process of operationalizing general goals, ultimately also in terms of achievement test items. Developing achievement tests is the last step in this deductive process.

There are many obstacles to apply this model: resistance against national standards and “state pedagogy”, incomplete knowledge about instrumental relationships, lack of vertical coordination between the central administration, intermediary organizations and schools, technical problems in getting from general goals to more operational formulations, and resistance by schools against externally developed guidelines and programs. Finally, the linear sequence from general goals to test items implies that didactic specification precedes evaluative specification and this is probably less efficient (see De Groot’s argumentation in favour of the opposite position in which evaluative specification precedes didactic specification).

Retroactive Planning, Combined with High Stakes Accountability

A less demanding type of planning than synoptic planning is the practice of using evaluative information as a basis for corrective or improvement-oriented action, sometimes indicated as “retroactive planning” (Scheerens et al. 2003). In that case planning is likely to have a more “step by step”, incremental orientation, and “goals” or expectations get the function of standards for interpreting evaluative information. The discrepancy between actual achievement and expectations creates the dynamics that could eventually lead to more effectiveness. In cybernetics the cycle of assessment, feedback and corrective action is one of the central principles.

Evaluation—feedback—corrective action and learning cycles comprise four phases:

- measurement and assessment of performance;
- evaluative interpretation based on “given” or newly created norms;
- communication or feedback of this information to units that have the capacity to take corrective action;
- actual and sustained use (learning) of this information to improve organizational performance.

This model resembles approaching alignment by given precedence to what De Groot indicates as “evaluative specification”.

When evaluative specification proceeds curriculum and didactic specification, it could also be seen as “taking the lead” in a more substantive way. Substantively processes of curriculum specification and test construction have much in common. This is particularly the case if we focus on learning tasks and assessable educational objectives. Admittedly, curriculum design has a broader scope, in also needing to address the choice and development of means (teaching strategies, classroom organization, and application of educational resources) apart from just having to select and ultimately implement subject matter based *content*. When recognizing the thoroughness of achievement test development one might wonder why a parallel process of specification and a parallel intermediary structure would be required in the didactic specification column. All that would be required might be an additional unit in a test development agency which proposes evidence based teaching strategies in relationship to content elements and educational objectives. Next, specific technical issues should be met.

Firstly, construction teams should have multi-disciplinary expertise with subject matter specialists in the lead, supported by didactical experts and test development experts.

Secondly, tests should be curriculum valid, relative to national standards and criterion referenced.

Thirdly, “half products” of test development like test frameworks and the specification of sub-domains should be made publicly available; for example to advice textbook writers.

Fourthly, calibrated item banks should be publicly available as well, in order to allow targeted test preparation by schools (van der Linden 1985).

Finally, moderate or high stakes accountability regimes would give schools a motivational impulse to align their teaching with educational objectives, standards and tests.

Particularly the fourth condition, calibrated item banks, allowing for legitimate test preparation would, in principle, be a strong step forward in attaining content alignment between what is taught and tested.

Ways to Empirically Assess OTL

Empirical procedures to measure OTL vary according to the *scope of the OTL definition*, the *data source*, the *level of the curricular unit envisaged*, and whether *exposure* or *alignment* is measured.

Scope of the OTL Definition

The basic definition of OTL refers to educational content. Further elaboration of this basic orientation considers qualitatively different cognitive operations in

association with each content element, often also expressing accumulating complexity (see the example from PISA 2012, presented in an earlier section). A next step in enlarging the scope is to add an indication of the time students were exposed to the specific content elements. Sometimes the theoretical option to include *quality of deliverance* to the OTL rating is considered as well. This option will be disregarded here as it is seen as stretching the OTL definition to a degree that it approaches a multi-dimensional measure of overall instructional quality.

Data source

OTL measures may be based on teacher judgments or student judgments. A third option is to consider unusual scoring patterns as instances of OTL differences.

The level of the Curricular Unit Envisaged

Curricular sub-domains, more specific topics, or test-items represent different levels at which OTL may be assessed.

Exposure or Alignment

The independent variable in assessing the impact of OTL on achievement can be a measure of exposure (has this content element been taught or not, or with a certain frequency) or an alignment measure. An example of an alignment measure as the independent variable is the correspondence between a measure of exposure and the contents of standards or assessment instruments. So in the latter case alignment is first assessed by means of content analyses methods, and then correlated with student achievement. An example is provided in the study by Polikoff and Porter (2014) which will be discussed in more detail in the next chapter.

Since basically all these dimensions on which OTL measure may differ can be crossed with one another, it follows that there is a broad range of ways to empirically assess OTL. This diversity could be seen as a problem when the objective would be to conduct meta-analyses of OTL effectiveness research.

Conclusion

What seemed like a relatively simple concept, at second sight, OTL proves to be rather complex. From the perspective of curriculum research, but also in fairly recent systemic modeling of educational effectiveness research (Scheerens 2016), OTL is part of a range of alignment issues, usually involving national standards, prescriptive formulations at intermediary level, like school standards and test frameworks, actual teaching and ultimate achievement measurement. Operational definition and measurement of OTL is also complex, in the sense that many options are possible, depending on the scope of the operational measure, measurement source, the curricular unit used to define OTL and the question whether OTL is operationalized as exposure or alignment.

In the final sections of the chapter, optimizing OTL was connected to the way educational systems are organized, particularly with respect to those facets of the system created to play a role in curriculum and test development. A preliminary conclusion was that alignment is an ideal of strong matching and coupling, projected in an actual organizational context that is usually “loosely coupled”.

The option to give precedence to what De Groot calls “evaluative operationalization” puts the spotlight on test-preparation, which in its turn opens up the question about legitimate and dysfunctional applications (teaching to the test). We shall turn back to all these issues in the last chapter of the report, in which we shall also formulate recommendations for educational practice and policy. But this will be done after we have taken a thorough look at the research evidence, concentrating on OTL effects on student achievement, which is the main issue of this study.

References

- De Groot, A. D. (1986). *Begrip van evalueren*. 's-Gravenhage: Vuga.
- Holcombe, R. W. (2011). *Can the 10th Grade Mathematics MCAS Test be Gamed? Opportunities for Score Inflation*. Unpublished manuscript. Harvard University, Cambridge, Massachusetts.
- Muys, D., Creemers, B., Kyriakides, L., Van der Werf, G., Timperley, H., & Earl, L. (2014). Teaching Effectiveness. A state of the art review. *School Effectiveness and School Improvement*, 25, 231–257.
- OECD. (2014). *PISA 2012 Results, Volume I, What students know and can do. Student performance in mathematics, reading and science*. Paris: OECD Publishing.
- Pelgrum, W. J. (1989). *Educational assessment. Monitoring, evaluation and the curriculum*. Enschede: Febo Printers.
- 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., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new U.S. intended curriculum. *Educational Researcher*, 40(3), 103–116.
- Scheerens, J., Glas, C. and Thomas S. M. (2003) *Educational evaluation, assessment and monitoring. A systemic approach*. Lisse: Swets & Zeitlinger.

- Scheerens, J. (2014). School, teaching, and system effectiveness: some comments on three state-of-the-art reviews. *School Effectiveness and School Improvement*, 25(2), 282–290. ISSN 0924-3453.
- Scheerens, J. (2016). *Educational Effectiveness and Ineffectiveness. A critical review of the knowledge base*. Dordrecht, Heidelberg, London; New York: Springer. <http://www.springer.com/gp/book/9789401774574>
- van der Linden, W. J. (1985). *Van standaardtest tot itembank. Inaugural Lecture*. Enschede: University of Twente.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21, 1–19.