Teacher literacy in formative assessment; the RTTI-system

Marinka Drost, Petra Verra and Jaap Scheerens

Abstract

Formative assessment is increasingly being recognized as a strong strategy of evaluation centered eduational improvement at school and classroom level. On the basis of an analysis of the key tasks that are involved we discuss the basic skills that teachers should possess to carry out formative assessments and conclude that these are quite ambitious.One of the solutions to improve teacher literacy in formative assessment isto give the teachers supporting methods and tools. In the main body of the chapter we describe an example of such a program. The RTTI approach is based on a meta-language about learning using practical web applications. It is presented as a structured instrument that facilitates the implemtation of scucessful formative assessment and improvement of student learning. In a way the approach compensates for some of the challenges that teachers meet. In a final discussion we return to the issue of basic requirements for professional development in function of formative assessment literacy.

Key processes of evaluation and formative assessment

All forms of evaluation consist of systematic information gathering and making some kind of judgment on the basis of this information. A further expectation is that this “valued information” is used for decisions on the day-to-day running of education systems or for more involving decisions on the revision and change of the system. The term “monitoring” is to be seen as a further qualification of evaluation, stressing the association with ongoing information gathering as a basis for management decisions, a reliance on administrative data and a stronger pre-occupation with description than with “valuing” (Scheerens, 2004)

When evaluative information is fed back to the units concerned, this can be an important basis for corrective action and improvement. The evaluation-feedback-action sequence is a central mechanism for all kinds of learning processes, including so called “organizational learning” and “formative assessment. The idea of learning from evaluation is central in the concept of *formative* evaluation, which is usually included in schemes for design and development in education.

As stated before, evaluating means judging the value of an object, and evaluation in the sense of a particular type of disciplined inquiry emphasizes that this “judging” and “valuing” is based on some kind of systematic information gathering approach.

In the case where this systematic information gathering is formalized according to the criteria for social scientific inquiry the term *evaluation research* is appropriate. A third major component of evaluation, next to the valuing aspect and the systematic approach to information gathering, is the applied context: evaluation results are expected to be used by relevant audiences. Again there is a prototype situation, often related to policy-evaluation, where evaluation results are expected to shape, or at least have a certain impact on, policy *decisions*.

In the evaluation literature authors vary in their emphasis of each of these three basic components: valuing, systematic inquiry and use for decision making. In all types of definitions where goal attainment is placed central, the value aspect is prominent (since whether or not program goals are attained provides the basis for judging it as either successful or unsuccessful). Thus Tyler defines evaluation as “*The process of determining to what extent educational objectives are*

actually being realized" (Tyler, cited by Nevo, 1995, p. 10).
In summary key processes of all types of evaluation are:

- (1) data collection by means of curriculum valid tests –
- (2) interpretation of test results by means of benchmarks for desired achievement levels –
- (3) diagnostic and formative interpretation of results – and applying these in providing feedback and developing feedforward

In formative assessment there is a particular emphasis on phase three; actually this phase is the defining characteristic in what we mean by "formative". At the same time data collection and interpretation of results may be problematic, as they touch upon scientific criteria of valid and reliable measurement and causal association between malleable processes and outcomes. In order to document this further we take a closer look at the three phases.

Phase 1) data collection
Phase 1 might involve the development of tests by teachers, which requires know-how on how to construct tests and test items.
But even if the tests are "given" this phase requires basic knowledge about psychometrics (what is a p-value, what is an average test score, what is a correlation coefficient, what is validity, what is reliability).
Phase 2) Setting norms, standards and benchmarks
Phase 2 assumes knowledge about norms, standards and benchmarks as a basis for interpretation of test results (e.g. comparing to last year's results, comparing to national norm scores)
Phase 3) evaluative interpretation
Phase 3 involves drawing evaluative conclusions for individual students and groups of students (diagnostic use) but also the development of conjectures on how remediation and improvement could be given shape; this last phase must be seen as very challenging.

Teacher competencies required to carry out these key processes of formative assessment

Evaluation studies (Visscher and Ehren, 2011) have pointed out that teachers mostly lack basic skills and knowledge on how to interpret quantitative test outcomes and to construct their own tests. To a smaller extent this also applies to the terminology on norms, standards and benchmarks, and how to use these constructs in practice. But also phase 3 is a particularly demanding task. It calls upon:

- Diagnostic skills – capacity to analyze mistakes- knowledge about the general capacity of individual students
- Knowledge on how to go "from diagnosis to therapy", in this case the planning and adaptation of the next period of instruction (sometimes indicated as feedforward)
- In a technical sense this would mean that teachers should have research skills; in terms of teacher professionalism this could be expressed as teachers possessing "pedagogical content knowledge".

To illustrate this further let's consider the technical and professional requirements that are involved. Assuming the availability of test or assessment data which reflect student outcomes and progress students make in realizing these outcomes, the next challenge is to attribute these outcomes to "educational treatments", in other words the methods that teachers have used, the coverage of subject matter content, the atmosphere during lessons etc. When these tasks are framed in a research context, they are seen as providing major methodological challenges. A big question is how teachers, not formally trained as researchers, might be expected to deal with such issues.
After this is solved, and strong and weak performance of students is credibly attributed to facets of teaching and instruction, in other words a "diagnosis" has been made, the next question is how to make the step from diagnosis to "therapy".

Three approaches can be considered:

- Calling on the experience and professional skills of teachers and trust their clinical judgment on how to proceed; possibly assisted by colleagues
- Finding out "what works" in the practical situation itself, by carrying out experiments.
- Making use of "evidence based' solutions known from educational research.

These demands on teacher's evaluative competencies are sometimes subsumed under the title "teachers as researchers". The tenability of this perspective should be considered with some caution.
Carrying out experiments and knowing how to make use of the research literature are very ambitious.
From our perspective, having often realized that we as professional researchers have great difficulty in carrying out experiments in education, and knowing how complex, diversified, contested the educational effectiveness knowledge base is, these ambitions may seem unrealistic, but we see two possible directions for solutions:

- We succeed in the development of a "grass root, simplified research methodology for teachers. An example is the "data team method" by Schildkamp and colleagues (Schildkamp, Karbutzki and Van Hoof, 2014).
- We provide tools for schools that guides test based planning of teaching by means of multi- media packages.

In the main body of this chapter we present the RTTI approach as an example of a multi-media package that facilitates formative assessment and supports teachers in meeting the key tasks, which were discussed in this section, in a responsible way.

Description of RTTI

The RTTI approach and its foundations

In a very general sense RTTI can be seen as a method to improve teaching and learning and enhance student performance. As such RTTI can be placed within the research tradition of educational effectiveness and school improvement. The RTTI approach is associated with one of the most successful strategies, namely "evaluation driven reform strategies"
Such strategies have the following characteristics:

- *Focus on student achievement outcomes*
- *Interventions that directly impinge on student learning*
- *Assessment driven*
- *Alignment of assessment and planning*
- *Cyclic application of assessment (diagnosis, feedback)*
- *School organizational conditions in a secondary, supportive role*

Evaluation and feedback can be seen as driving improvement at school and classroom level. Implied facets are clarity of purpose through standards, examination syllabi etc., verification of what students have learned, identification of strengths and weaknesses in content and skills that are mastered, feeding back and diagnosis of outcome patterns, systematic consideration of remedial strategies and setting concrete goals for improvement at student, classroom and school level, in cooperation with other teachers, school principals and eventual support staff. This latter characteristic could make evaluation/feedback/systematic corrective action the core of task related professional development and teacher cooperation.

Zooming in on an evaluation centered approach to educational improvement

A convincing rationale

Educational effectiveness research has a peculiar relationship with educational practice. On the one hand "what works" as established in research seems directly applicable. On the other hand the research approach which concentrates on

separate factors, of necessity does not capture the full complexity of educational practice. An evaluation centered approach can be theoretically positioned in the context of cybernetics. Cybernetics is described as the transdisciplinary approach for exploring regulatory systems. The key mechanism consists of a sequence of evaluation, feedback and corrective action; which can be thought of as a cycle. The practice of using evaluative information on organizational functioning as a basis for corrective or improvement-oriented action is likely to have a "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 theory" the evaluation centered approach has several strong points:

- The clear focus on educational achievement guarantees remaining close to the core "production process" in education, namely that of a learning student.
- The possibility to align assessments, including formative assessments, to examinations and summative high stakes tests, can be seen as a strong mechanism to enhance "opportunity to learn" (as a good match between what is taught and what is tested).
- Feedback based on assessment results is an important impetus for both student learning and teaching, with cognitive and motivational implications. Feedback refers to reporting back performance results to the producers, i.e. students and teachers. Information on certain deficits can be used in subsequent didactic planning, and research has shown that such "instrumental feedback" stimulates student learning (Kluger and DeNissi, 1996)
- When taking control theory to the letter a good evaluation mechanism is sufficient to allow effective organizational functioning; particularly in the case of professional organizations like schools everything else could remain "deregulated".
- Finally, system level accountability policies can also be seen as part of an evaluation centered strategy for educational effectiveness. Although this position is not without being contested, external accountability can be a stimulating condition for "internal accountability" (school self-evaluation and formative assessment) (Carnoy et al., 2003).

Mixed empirical evidence

Review studies and meta-analyses concerning evaluation, assessment en feedback are, among others, those by Black & Wiliam (1998), Hattie & Timperly (2007), Shute (2008), Hattie (2009) and Wiliam (2011 Fuchs & Fuchs, 1985, Kim, 2005, Hattie, 2009, Burns & Symington, 2002, Bangert et al., 1991, Kluger and DeNisi, 1996, Faber and Visscher, 2014, Hendriks et al., 2014. Scheerens, (2016) makes a distinction among meta –analyses of studies about *formative evaluation*, *assessment and testing at classroom level* and studies about *feedback*. The difference between the categories "formative evaluation" and "assessment and testing" is that for the latter category of studies no explicit indication was given whether tests were applied as formative or summative.

The results show sizeable differences between the various meta-analyses, for each concept. Among the three types of evaluation that were discussed, the results on feedback are particularly high ranging from .51 to 1.10. Results on formative evaluations vary between .06 and .70 and those about assessment and testing at classroom level between .10 and .39. Scheerens' results also indicacte that effect sizes found in more recent meta-analyses and review studies tend to be considerably lower than those obtained from older studies. When comparing these outcomes to results of meta-analyses on *instructiontime* and *school leadership* the results on evaluation, testing and feedback were considerably higher (Scheerens, 2016, ibid). The overall higher effect size for feedback is probably partly due to the fact that most of the results on feedback variables were computed in micro level studies at the classroom level.

High potential, sub optimal implementation

Although the general picture of the evidence cited in the previous section compares favorably to the results of meta-analyses on other often studied effectiveness enhancing variables, they may still seem less than optimal, given the theoretical advantages presented earlier (Scheerens & Bosker, 1997, Scheerens et al., 2007 and Hendriks et al., 2014). For

various reasons application an implementation of evaluation centered approaches in schools are often constrained by a number of cultural, political, technical and practical issues:

- *Cultural resistances.* In education, an output oriented perspective has traditionally met with a lot of suspicion. When combined with educational testing it evoked outright resistance. As predicted by organization theory of the school as a "professional bureaucracy" highly trained professionals do not like external evaluation of their work (Mintzberg, 1979). Even when evaluations are formative student evaluations it is hard to get rid of this bad image in schools. Although in recent decades this situation has changed, the process of including evaluation and evaluation centered approaches in schools is rather slow in coming.
- *"Ownership issues".*In many educational systems school and teacher autonomy has been strongly stimulated for the last three decades. The instruments for evaluation and assessment are often developed externally to the school. This brings the danger of these literary remaining "Fremdkoerper" (alien elements) in school life.
- *Incomplete application.* There are many examples in which schools have gone to considerable trouble to collect evaluative data, and never use it. Tables and reports die a quiet death in cupboards, and are not even discussed among staff, let alone used to improve student learning. The phases of feedback, diagnostic interpretation of data and providing input for subsequent didactic planning are crucial in making evaluations pay-off.
- *Association with administrative burden and bureaucracy.* Collecting data for evaluation purposes takes time, and keeping records on the basis of these data perhaps even more. Internationally there are unfortunate examples where schools and teachers are compelled to make many extra hours to carry out such tasks, even in situations where the relevance for improving teaching and learning is unclear.

Considerations for successful application

The constraints that were mentioned in the section above can be taken as challenges to improve the situation. The following solutions should be considered:

- *Rethink the connectivity between formative and summative assessment.* Concentrating on formative assessment seems to be the easy way out of the problems listed above. Still the pay-off for enhancing educational effectiveness would be greater if formative assessment would be well-aligned with high stakes summative tests and examination.
- *Maximize teacher autonomy and "ownership".* Depending on the leeway that schools have to buy –in material and external support, commercially available products that are bought have the advantage of literary becoming "owned" by the school. Next packages might provide ample opportunities for schools and teachers to use material in the most relevant way, dedicated to the particular situation; for example teachers might be stimulated and supported in developing their own tests.
- *Stimulate and support a "cyclic" application of testing.* A major challenge is to get schools to go beyond mere data collection and get to the actual application and use of test results. In the Dutch context such an approach fits perfectly with the policy to stimulate "achievement oriented work" (Dutch: *opbrengstgericht werken*). This is not a small thing to accomplish and requires support in the form of teacher professional development and detailed practical guidelines.
- *User-friendly ICT applications.* Such applications are important to put across material and support for key processes of cyclic test-application, but also to facilitate student administration and record keeping.

Publications about educational effectiveness and school improvement tend to breathe a positive message of recipes for success. An analysis of the research evidence is more of a sobering experience. Even though there is considerable agreement as to which malleable conditions of schooling and teaching matter, the quantitative evidence on "how much" these conditions work is much more divided. Moreover, effect sizes are often quite small, when compared to established standards and often smaller than those of "nonmalleable" given conditions, such as student aptitude, socio economic status and aggregates of these variables at school and classroom levels. Moreover, improvement in the "process

conditions" of schooling often requires considerable time and effort to show up in relatively small increments of progress in student achievement. In other words one often needs to invest a kilogram in process improvement to obtain an ounce of success in output. What the research literature does show is that some conditions work better than others. Tinkering with school organizations and the umpteenth fashion in school leadership approach is not going to have much effect, unless there is a clear connection with strategies that impinge on the primary process of student learning. Improvement strategies closely connected to student learning and the assessment of learning effects have the potential to be not just effective but also efficient. The RTTI approach, as further described and specified in this article is a candidate for an effective and "lean" approach to enhance student learning. In the final section we will return to the issue of teacher formation. This final discussion will treat two issues; the first is to summarize how a structured method like RTTI helps teachers to compensate for breaches in their training and preparedness and the second is to return to basic training that is still required even with the degree of 'scaffolding' a method like RTTI offers.

Description of the RTTI approach to enhance student learning

When we introduced RTTI as an evaluation centered approach by teachers at school and classroom level, the following observations were made about the current educational situation in the Netherlands:

- Teachers sometimes lacked testing competencies and did not have the required know-how on how to construct valid and reliable tests and test items themselves. For existing tests, their basic knowledge about psychometrics appeared to be often insufficient (e.g. in order to deal with questions like: what is a p-value, what is validity, what is reliability). There was not always enough attention for teacher training to construct and analyze tests (Sanders, 2013). Furthermore, there is hardly any practically useful literature on test construction as the available publications tend to focus mostly on psychometrics instead of being applicable for teaching practice. Teachers frequently use tests that are embedded in teaching methods and textbooks, and because of their widespread use it is assumed that these tests are of good quality, which is not always the case. The content validity of these method-bound tests is generally excellent, but the construct validity and the way the test is constructed is not. Alignment between the various tests is often inadequate: for many subjects in secondary schools there is a gap between the undergraduate and graduate levels, as tests and content - especially in the lower years - are often focused on lower order thinking and unsuitable for formative evaluation.
- Regularly, data from external tests are kept by school management, and are not shared with teachers. Also, in situations where test data is made available to teachers, not all teachers can interpret and use that information. Great differences exist between research skills of teachers to carry out further research in response to these data (Datnow et al., 2012). A considerable number of teachers is still insufficiently capable of going from monitoring to diagnosis and from diagnosis to therapy (ibid, 2012).
- The learning objectives are not always clearly indicated by teachers. What does it mean if a student has to study the third chapter for a test? Some students think that they must literally memorize the text of chapter 3, others that they must practice a lot with the given assignments, still others that things will be fine because they participated in the lessons, so they do not need much preparation. There is a lack of specification of goals and content (Folmer, 2017, p.87). Because of this, it is not always clear to students what is expected in terms of preparation. Furthermore, feedback from teachers is not always effective.
- In addition to providing a basis for grading, tests are increasingly being used to provide information about the learning process to teachers and students. Teachers are in principle positive towards formative evaluation, provided it does not give them too many extra administrative responsibilities.
- The culture and organizational arrangements at some schools does not sufficiently support formative evaluation (Datnow et al., 2012; Oprins & Andriessen, 2002). This is partly due to the lack of a quality cycle. Elements of the quality cycle (formulating goals, building in evaluation moments, feedback and accountability discussions), ensure that quality as such is discussed. Teachers already experience their work pressure as high, so there is little time

to think about quality improvement. Dealing with the normal daily issues prevents reflection. In addition to facilitating structural arrangements, the quality *culture* at a school is an important success factor, meaning a culture where teachers actively deal with information, address each other, work together and have opportunities to learn from each other (Onderwijsinspectie, 2015). Structure and culture can reinforce each other so that there is sufficient basis and time for and focus on formative evaluation and evaluative interpretation (Datnow et al., 2012).

It is not so much that teachers are not eager to improve their evaluation practices, but they do not know how, and the system does not support them in this (Lyons, 1998). Established routines are difficult to change and renew in a more fundamental way (Onstenk, 2004) Moreover, the immediate educational situation gives little time for reflection. Teachers make split second decisions that result in immediate action (Eraut, 1994). To carry out the key processes of formative assessment, it is therefore necessary to supplement teacher competencies and to have the school culture and structure support these (Onderwijsinspectie, 2015). This requires an approach with an effective meta-language about learning, so that there is an unambiguous terminology about learning at school that can be used for all subjects, all grades and school types. This approach needs to be supported with practical methods and tools for formative evaluation, ideally through web-based applications, in which concrete feedback on the test construction from psychometric values is pre-structured for teachers. It also requires addressing the need for an easy to incorporate quality cycle in the existing school structure combined with a quality-oriented culture.

A meta-language for formative evaluation

The cognitive classification RTTI, a meta-language about learning to enhance student learning, is developed from educational teaching practices and from existing systematic classifications of question types, also called 'taxonomies'. One of the best-known taxonomies is (revised) Bloom. We first tried to use this taxonomy as a meta-language but using the 19 categories of revised Bloom is not efficient and using the six main categories is not effective for formative evaluation as they do not efficiently map the learning process to provide effective targeted feedback. In addition, the validity of the continuous learning path is not secured as the six main categories do not provide sufficient grip on the learning objectives, as is discussed below.

Chapter 17 of the (revised) Bloom Handbook mentions several problems which, once solved, would make the framework more useful (Anderson & Krathwohl, 2001, p. 295). In the current chapter we go beyond these solutions of revised Bloom, and draw attention to the following issues

- *The process of integrating curriculum, instruction, and assessment*
- *The usefulness of the framework to students*
- *The relationship of the framework to a widely accepted, meaningful, and useful theory of learning*

These challenges have been addressed with the development of RTTI and the related tools, together with the aforementioned starting point for carrying out the key processes of formative assessment.

From Revised Bloom to RTTI

The first main category of the taxonomy of Revised Bloom is *Remember,* which is similar to the definition Remember (R) within RTTI.

Remember questions (R) are intended to test students' recognition, retention, and reproduction of crucial and relevant factual information. The subject matter for such questions is explicitly indicated or marked. When answering R questions students are not expected to add anything to the content.

The second main category of Revised Bloom, Understand, is in practice perceived by teachers and students as diffuse and not efficient enough for formative evaluation. For an efficient formative evaluation, the seven skills of Understand,

have been subdivided by into Training (T1), Transfer (T2) and Insight-Innovation (I). Revised Bloom's Exemplifying and Classifying belong to RTTI's Training. Interpreting, Comparing and Explaining belong to RTTI's Transfer. Summarizing and Inferring belong to RTTI's Insight.

For the third main category of Revised Bloom, Apply, it appears that a large part of the test questions and content falls in this category. This is not only the case for primary and secondary education, but also for higher education. In order to provide students with targeted feedback, and to organize education responsibly and effectively (curriculum alignment), it is necessary to make a distinction within Apply. RTTI explicitly distinguishes two application levels through Training (T1) and Transfer (T2). T1 and T2 from RTTI show similarities with the two underlying skills from the main application category of Bloom: performing a procedure in a routine task/in a known context (Apply - Executing) versus executing in a new task situation/in an unfamiliar context (Apply - Implementing). In the learning process of the student there is an important difference between those two application levels. In practice, there are students who are perfectly capable of solving the trained application questions (T1) with simple and indisputable answers, but with transfer-oriented application questions (T2) they have to answer based on the changing context. For people with impairments in the prefrontal cortex, we see that for transfer-oriented questions, other areas of the brain are active than for the veridical decisions (Sousa, 2009). In addition to training-oriented T1 questions, it is also important to separately distinguish transfer-oriented T2 questions, because "Learning to generalize (transfer) is the core of good education. If our students can only reproduce what we have learned them, they cannot do much, because those exact circumstances will probably never occur again. Learning is only useful if the student can apply what he or she has learned outside the context of the learning process" (William & Leahy, 2018, 43).

For a training question (T1), the application is practiced, the source is recognizable, and the question has a similar degree of difficulty as the one in the workbook. Students recognize the variables in the new source, and the issues that they have to pay attention to are recognizable and matched to practice during lessons. As mentioned earlier, Exemplifying and Classifying from Revised Bloom belong to T1. Exemplifying occurs when a student must select or produce a specific example or instance of a general concept or principle, e.g. 'give examples of various artistic painting styles' (Anderson, et.al., 2001). Classifying is a complementary process to exemplifying and occurs when a student recognizes that something belongs to a certain category (ibid, 2001).

Training-oriented questions (T1) aim to assess the application of the given subject matter in a known situation, comparable with the one that was trained with. T1 questions determine whether the student has mastered learned step-by-step procedures, methods or 'recipes', similar to the situations previously trained with. It is therefore important that T1 questions do not prompt the discovery of new information. As such, these questions should not assess exceptions to a learned rule or be of 'red herring' in nature.

For a transfer application question (T2), the new source is different than the one students have trained with, and this must therefore first be 'cracked'. Furthermore, a T2 question cannot be addressed directly with a step-by-step plan with which students have practiced, instead, the T2 question can be solved by putting together and combining practiced solution steps. As mentioned earlier, Interpreting, Comparing and Explaining of revised Bloom belong to T2. Interpreting occurs when a student is able to convert information from one representational form to another. To increase the probability that interpreting is assessed, the information included in the assessment task must be new, not encounter during instruction (ibid, 2001). Comparing involves detecting similarities, finding one-to-one correspondences between elements and patterns in one object, event or idea and those in another object, event or idea. In comparing, when given new information, a student detects correspondences with more familiar knowledge (ibid, 2001). Explaining occurs when a student is able to construct and use a cause-and-effect model of a system, or find explanations for a symptom, or troubleshooting to give a diagnosis (ibid, 2001).

Transfer-oriented questions (T2) assess the application of the subject matter to a new situation, which requires a transfer. These questions assess student recollection, application, and transfer of learned materials into new contexts. Namely, it aims to determine a student's ability to select the correct variables / data, determine which procedure is the most suitable, and to combine parts of plans or procedures to solve T2 questions.

Revised Bloom's three main categories Analyze, Evaluate and Create, have all been merged into the RTTI category Insight-Innovation. Revised Bloom also reports that these three often go hand in hand with metacognitive knowledge (p.239). The three main skills are rarely taught in school practice, including in higher education, because normal emphasis is on other skills. Many school topics fall into three cells (Remember, Understand and Apply) of the Taxonomy Table (ibid, 2001). The main reason for combining these as RTTI's Insight-Innovation is because of efficiency.

A student can answer an Insight-Innovation question if he or she is able to construct the context or method independently, and to work innovatively, either inside or outside of the curriculum. The student must therefore be able to place and frame the source himself and must construct his own, untrained and not pre-structured solution strategy in order to arrive at a correct solution. As mentioned earlier, Summarizing and Inferring of Revised Bloom belong to RTTI's Insight-Innovation. Summarizing involves constructing a representation of the information (ibid, 2001). Inferring involves a pattern within a series of examples or instances, a student is able to abstract a concept, comparisons among instances within the context of the entire set (ibid, 2001).

Insight-Innovation questions (I) require the student to analyse and evaluate independently and systematically from various perspectives, the student himself has to create and construct both context and the procedures, whether or not outside the curriculum, that are necessary to come to a solution.

Below is an overview of how the six main categories and nineteen skills of Revised Bloom relate to RTTI.

RELATION RTTI – REVISED BLOOM	
R	➢ **Remember –** Recognizing; **Remember** – Recalling
T1	➢ **Understand** – Exemplifying; **Understand** - Classifying ➢ **Apply** – Executing
T2	➢ **Understand** – Interpreting; **Understand** – Comparing; **Understand** - Explaining ➢ **Apply** – Implementing
I	➢ **Understand** – Summarizing; **Understand** - Inferring ➢ **Analyze** – Differentiating; **Analyze** – Organizing; **Analyze** - Attributing ➢ **Evaluate** – Checking; **Evaluate** - Critiquing ➢ **Create** – Generating; **Create** – Planning; **Create** - Producing

Table 1: *Comparing RTTI to Revised Bloom*

The non-hierarchical learning processes

"In a taxonomy, the categories lie along a continuum." (Anderson, et al., 2001, 4) and describe a hierarchical classification system (Block, 1975). Learning is not always stacked, which is why RTTI is not defined as a taxonomy but is called a meta language about learning. R is not always the preliminary stage of I: sometimes students who answer the I-questions correctly would not score as well for R and/or T1 and/or T2 questions. Some students think difficult things are easy and easy things difficult (William & Leahy, 2018). A teacher has to discover the learning logic of a student. This is only possible if you map all cognitive levels, so from R to I, in order to gain insight into the learning process and into the talents of students (and classes) thus understanding the students' "learning entrance". Therefore, tests that are used for formative evaluation must include questions from R to I.

Data collection for formative assessment

The process of determining to what extent educational objectives are actually being realized (Tyler, cited by Nevo, 1995, p. 10)*,* starts with the development of assessment and data collection. *"We use the general term assessment to refer to all those activities undertaken by teachers – and by their students in assessing themselves – that provide information to be used as feedback to modify teaching and learning activities. Such assessments become formative assessment when the evidence is actually used to adapt the teaching to meet students needs" (Black & William, 1998, p.2).* This might involve the development of tests by teachers, which requires know-how on how to construct tests and test items. Test construction does not always get the attention that is required. The available literature on test construction is mostly written from a psychometric perspective and is not practical enough for the daily teaching practice of teachers. A teacher often has no opportunity to do pretests for example, and frequently he or she does not have the time and expertise to perform and interpret complex calculations. However, if teachers are expected to get more information from tests than just a grade, they will have to be able to conditionally check whether the basic quality of a test is sufficient. They will also have to be able to correct construction errors, perform basic analysis and interpretations in the psychometric field, and decide what steps can be taken to go from analysis to diagnosis to therapy.
In the Handbook RTTI (Drost & Verra, 2018a) various kinds of practical tools are offered to stimulate and support teachers in developing and checking their own tests, including practical directions to use material in the most relevant way, depending on the situation. A functional *Check and Design list* helps the teacher, before administering a test, to check the basic quality of the test questions. For basic teacher literacy in the field of RTTI as a meta language and psychometrics, an e-learning tool has been setup, and a supporting web-based tool for formative evaluation has been developed, called RTTI-online.

Formative use of assessments

Assessments are usually either formative or summative, but the label isn't nearly as important as what you will do with the information that has been collected to enhance students learning. Information collected from a summative test can be used in a formative way (Laveault & Allal, 2016). If the assessment is formative, a certain degree of generality must be built into the learning objectives and success criteria in order to promote transfer (Arter & McTigh, 2001 in William & Leahy, 2018). Tests are used less and less to solely measure a student's performance in a grade. Instead, it is becoming increasingly common to also extract information in service of the further development of the student. Test results should provide information about the depth of learning, different levels of complexity, autonomy of learning, more or less guidance or support, and transfer of learning, to qualify for learning and / or novel situations (Laveault & Allal, 2016).
As far as the formative use of tests is concerned, teachers first need to find the cause for a (lagging) result. For such formative evaluation, the test should contain questions from R to I. For example, if a student fails on a T2 question, the cause may lie in a lack of knowledge on R, in insufficient practice with T1 questions, or -if the student does well on I- in the linguistic context of a T2-question. It goes without saying that an RTTI ratio must be chosen that fits the learning

objective, school type and year layer.
Beside that teachers can stimulate students based on their strengths, because RTTI explains the given grade. For example, for a student who scores very well on I but fails on R, to find logic in the material to be learned will be much more beneficial than just chopping the subject matter into pieces and repeating this.
In practice, however, not all tests contain questions from R to I. The example question constructions in the table below can be used by teachers to design additional questions for formative tests if a certain cognitive level is lacking.

RTTI	Example questions for constructing additional questions
R	Give the definition of; Give the step-by-step plan; Give the formula of; Give the enumeration of; Appoint; Point out; How / where / when / who (fact)?; call / translate 'basic knowledge'
T1	Calculate; Search; Group; Describe in your own words; Give an example of; Fill in; Allow learned characteristics to be used; Conjugated; Play after / Image off; "Search-find" question
T2	Argue what fits best; Show with ... that; Make an error analysis; Explain; Explain with ...; Show with ... on; Give cause and effect relationship / connection; Make combinations; Compare; Declare; Observe; Appreciate; Coordinate
I	Construct a hypothesis; Develop / design / shovel; Analyze: Give a new argument for; Give a relevant summary; Critical assessment; Derive (similarities / differences); Predict; Monitor and conclude; Design a research design; Give advice; Evaluate; Coach; Proof

Table 2: RTTI *Example questions*

Evaluative, diagnostic and formative interpretation of results

If teachers have a formative test available, this does not mean that they can also evaluate formatively. The basis for interpretation of test results assumes not only knowledge about norms, standards and of desired achievement levels, but also knowledge about the diagnostic and formative interpretation of results – and applying these in providing feedback and actions for individual students and groups of students (diagnostic use). This also includes the development of conjectures on how remediation and improvement could be shaped. This phase is very challenging for most teachers. Teachers see formative evaluation not only as complicated, but also often as an extra task in addition to their teaching activities, rather than as part of it (Datnow et al., 2012). Support with concrete feedback, feedforward and examples of remediating actions is crucial for teachers. Teachers are not formally trained as researchers, yet are expected to deal with these issues. Training this aspect is very important because of a focus on educational achievement and strategy that impacts the primary process of student learning. This can best take place as training on the job, making the shift away from a traditional focus on transfer of knowledge and skills, to developing understanding of, and in, professional practice (Dall'Alba & Sandberg, 2006). RTTI-online contains important elements to support teachers in their formative evaluation.

User-friendly ICT application
The web application RTTI-online provides support for the systematic information gathering, analysis, diagnosis and actions, tailor-made feedback and feedforward for both teacher, student and school management. RTTI-online can be used for any assessment, at any time, a self-made test, a national exam or an assessment provided with the teaching method.

1. It helps teachers properly balance the assessment regarding the different learning categories R, T1, T2 and I.
2. It automatically analyses the results to monitor and check quality.
3. It proposes targeted feedback and feedforward for teachers and students.
4. It provides formative evaluation.

The automatic reflection tool QA (Quality Analysis) for the teacher (a formative evaluation of the test) consists of the

following four steps.

Step 1) Monitoring curriculum alignment

When teachers first mark their tests with RTTI, they immediately ask about the optimal RTTI ratio of their test. This begs the question about the place of that test in the continuous learning line. A continuous learning line indicates how students, for a given course during their school career, progress from their starting point to the exam level. It is a continuous program in which, as a department, there has been deliberate thinking about five important aspects that ensure good substantive alignment: What are our learning objectives, what material is offered, at what time, in what way, and why (content validity). In addition to this substantive alignment, the continuous learning line also requires adjustment to which cognitive level the substance should be treated and in what RTTI ratio (construct validity). This prevents the cognitive leap for students to be too big (which would demotivate students) or might cause the average result of a class to suddenly decrease sharply. This is why for each year layer the learning objectives / topics are discussed, including in which RTTI-ratio these are taught and tested.

Backward mapping

The first step in the Quality Analysis of RTTI-online is the check on the alignment of the test with the continuous learning line. When there is too much deviation between the RTTI ratio in the current keys and the optimum ratio, the existing keys are gradually adjusted. Teachers need time and opportunity for these adaptations: the teacher cannot radically change the corresponding course sequence to the required RTTI ratio at once, and students also must be able to get used to the new way of problem definition and curriculum and need to get the chance to acquire the necessary strategies.
By making the learning objectives and corresponding tests transparent and concrete prior to each lesson cycle, the focus of all teachers is placed on the learning objectives to be achieved and the teachers' lessons become more efficient and effective (Davies, et al., 2016). It is important for the motivation of students that it is also clear to them what the learning objectives of the lesson cycle are (Laveault & Allal, 2016) and that teachers align the assignments and the homework with those learning goals. In practice, it does happen that homework consists of 90% T1 questions, while the learning goal for T1 is only 40%. And this could be one of the possible causes to explain low student motivation for doing homework.

Step 2) Monitoring the quality of test questions

The second step in the QA analysis is monitoring the quality of the test questions. In order to analyze the quality of the test questions, the determining value of a question is examined, for which the Rit[1] and Rir[2] values are often used. For many teachers, this is quite difficult to calculate and interpret. The p-value, the percentage of the candidates who answered a question correctly, is often used for judging the quality of the test questions. When a teacher works with RTTI he or she can only use (yet less reliable) p-values as a practical alternative to monitor the quality of the test questions. The p-values of the T2 questions and I-questions for example, for a first class of secondary education are expected to be lower than those of the R-questions and T1-questions. After all, there are normally fewer students in this class who will answer T2 and I questions properly, compared to R and T1 questions. In a fifth grade of secondary education for example it is expected that relatively more students will answer the T2 and I questions correctly than in the first grade, so the p values for those questions are expected to be relatively higher.
In order to guarantee the quality of the test questions, RTTI-online automatically signals when an RTTI-encoded question falls outside the expected value, for the relevant type of education. All questions that fall outside the expected value are shown to the teacher for further analysis:

- The cause could be a construction error, in which case the question needs to be checked for possible removing or improvement before use in a subsequent test.

[1] Rit value:Rest item test value reflects the perfomance of the item versus the test as a whole
[2] Rir value:Rest item rest value reflects the perfomance of the item versus the test minus the socre of the item

- The cause could also be an incorrect RTTI coding: the question appealed to a different cognitive level than previously intended, in which case the coding can be adjusted for a correct diagnosis.
- The cause could be didactical: the learning objective did not fit with the offered learning arrangement. The teacher can use didactic interventions at class level to bring the students to the desired level. Teacher colleagues whose classes perform better can also be linked, to enable the sharing of expertise.

Teachers check and secure autonomously the quality of their own tests and can possibly decide to remove test questions and / or improve them for future use. The teacher can also draw up repair programs for his didactics and learning arrangement.

Step 3) Didactic interventions

In step 3 of the Quality Analysis didactive interventions promising interventions will be proposed and visible for the relatively weakest cognitive level of the class. Firstly, recommendations for effective learning strategies that can be transferred to the class, for example through 'modelling/showing' or through the explicit use of the given learning strategies when discussing the homework. Secondly, the application gives concrete recommendations for the use of alternative teacher roles, and in addition the teacher can directly choose from more than 100 practical teaching methods that can be used in the various phases of the lesson. *This step supports teacher analyzes the errors that frequently occurred in order to design new activities that will help the students progress in their understanding of key scientific concepts that where studied* (Laveault & Allall, 2016, p.9).

Step 4) Pedagogical interventions

In step 4, the RTTI pattern of each student is shown, which gives insight into the learning process behind their grades. This RTTI pattern can be compared to the moving average of the student on his pattern by the teacher; this is idiographic research. The teacher gets a description (narrative) and possible cause for the RTTI pattern. "Feedback is most effective when it is the right kind (e.g. detailed and narrative, not graded)" (Andrade, 2013, 25). The most promising learning strategies that match the RTTI pattern are shown, so that the teacher not only receives an analysis and diagnosis, but also suggestions for directly targeted actions for the student for immediate use. The teacher can also view what the student himself has given as an action in his student account of RTTI-online and supplement this action if necessary. Finally, teachers and their professional colleagues can discuss (inserted) cases that have not yet started deploying actions and supplement these from the shared expertise. This can be disseminated to the entire team for further elaboration during student discussion.

RTTI quality cycle

The RTTI quality cycle supports the structure of formative evaluation in the school, in accordance with the recommendation of William & Leahy (2018) to create opportunities within the school to promote dialogue, to work on a coherent curriculum in which education, learning and testing seamlessly connect.

The RTTI quality cycle is run according to the number of teaching cycles in which the school year is set up at school; most schools work with 3 or 4 cycles per year. By analyzing an RTTI-marked test (built up according the optimal RTTI ratio) in each cycle, using the 4 steps of the QA, a good and reliable overview of the student's learning process for all subjects is created.

The RTTI quality cycle is facilitated by several specific events and activities. The numbering below follows the numbering in the adjacent figure of the RTTI quality cycle.

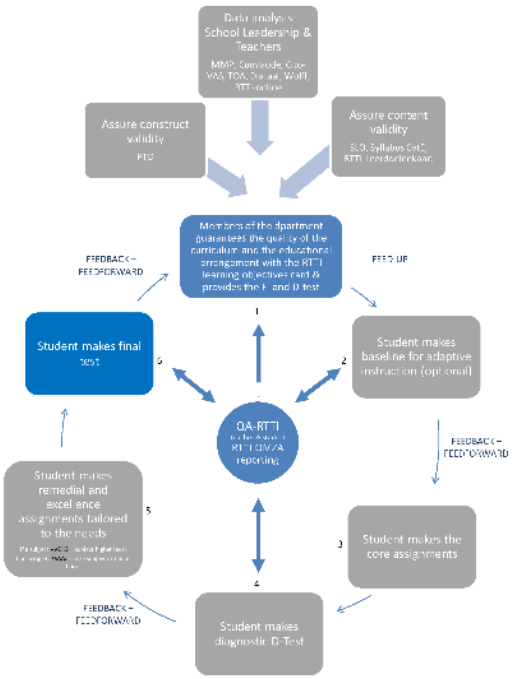

Figure 1, RTTI quality cyclus

1. Task teachers

The teachers supply and check the quality of the high-test assessments (V-tests) and diagnostic tests (D-tests) that will be taken in the coming teaching cycle and make RTTI study methods. William and Leahy (2018) recommend to first make the final test, so you have the knowledge and skills sharp and know want you want the students to be able to do at the end of the module or series of lessons. The tests are preferably taken simultaneously if a joint test or assessment is involved.

2. Zero measurement

The students basically make a baseline measurement in the first lesson of the teaching cycle. Teachers can use the diagnostic (D-) test for this or use a set of homework or assignments representative of the learning objective. Depending on the result, student will make the core assignments for those components that they do not master yet, and the excellence assignments for those components that they already master. With adaptive instruction, students who already fully mastered the subject matter can follow the course at a higher level or follow an accelerated program. The teacher can also use the baseline measurement option to accommodate students who like to first acquire the material through the core assignments and do a diagnostic test afterwards.

3. Core assignments

The core assignments are a limited selection of assignments from the entire curriculum. It is made up of the minimum program with the crucial and relevant components needed to achieve the learning objective. Dutch L.O.O.T. schools[3] have often already developed such core programs because they have reduced the curriculum to the core of the curriculum for their top-class athletes. For students who follow the regular program, this is an effective way to leave more time for customization (remediation), a deepening of the subject, another subject, or to be able to accelerate.

4. Diagnostic key (D key)

Approximately two weeks before the progress (V-)test, a diagnostic D-test will be taken, in accordance with the learning objectives and ratio on RTTI, so that the student, in addition to the subject-specific check, also gains insight into the comprehension of the different cognitive levels, as the V-key will come up for discussion. This can also be a representative part of the homework that is assessed by the teacher or the student. The teacher can also consider this test as a 'tool' if there is a need for an extra grade. The assessment of the D-test does not always have to be done by the teacher. The student himself or a fellow student can do this as well. Using the formative evaluation of the diagnostic test, the students can use specific actions to prepare themselves for the 'real' high-stake (V-) test.

RTTI-learning objectives card: Dutch

1. Reading Business texts 3 gymnasium		Periode:		Weighing (Final) test:
	R	T1	T2	I
RTTI-ratio	5%	55%	35%	5%
1.1g. You can summarize texts Toest bekijod link	Conceptual framework; a.o. • Main issues • Side issues • Summarize on indications • Summarize informatively • Summarize schematically Key issues are the parts of the text that are really needed to understand the story. Side issues are ... etc.	You can distinguish main and side issues in a text and recognize different types of summaries. https://www.youtube.com/watch?v=8qKpe-bsaF4	You can link the main issues of a text in a logical way https://www.youtube.com/watch?v=4F81-nWCJtw	You can make an adequate summary of a text for others that fits the communicative goal that needs to be achieved https://www.youtube.com/watch?v=8647-bGT49w
Support	Overhearing*	Signalwords	25-words-summary 5W-questions	OV-SLIM
Core assignment	B5-1a**	B5-1be, 2abc, 3, 8	B5-4, 5c, 6, 7b	B5-5ab, 7a
Remedial assignments	B5-R	B5-T1	B5-T2	B5-I
Excellention assignments	D5-1abc, 2a	D5-1de, 2bc, 4, 9	D5-5, 3, 6c, 7, 8b	D5-6ab, 8a, 9
Testitems database	4, 7	1-3, 8-12, 15-17, 22	5, 6, 13, 19-21	14, 18

*Leerstrategieën uit *Slimmer Leren met RTTI*, uitgeverij: Uitgeverijplus
**Voorbeeld uit methode *Kern Nederlands onderbouw*, uitgeverij: Staal & Roeland

Figure 2, RTTI learning objective card

5. Remedial and excellence assignments

The student knows (after the formative evaluation of the diagnostic test in RTTI-online) also his possible gaps and strengths. In the case of gaps, the student makes the remedial assignments, as indicated in the RTTI learning objectives card or the study guide. For the stronger cognitive levels, the student can get started with the excellence assignments. If all cognitive levels are already mastered, the student may also finish the course at a higher level and start working with corresponding assignments from the relevant learning objectives card.

[3] A secondary school where talented athletes have the opportunity to combine their top sport development with education.

6. Progress test (V-test)

Finally, the V-test is administered. After the tests have been checked, the students enter the scores in RTTI online and, with the aid of the QA, gain insight into their results and possible approach. "The most important thing is that we help students to learn without us." (William & Leahy, 2018, p.182). Based on that analysis, students themselves indicate which actions they will use to achieve better results in the immediate future. These actions of the students are combined with the possible actions from the QA of the teacher, after which they can be effectively combined. After this the next quality cycle starts.

A formative evaluation supported by RTTI-online makes it possible for teachers to carry out evaluation research, improve education, generate revenue for their own teaching practice through learning on the job and at the same time provide targeted development-oriented feedback to students.

Conclusion; RTTI and the development of literacy in formative assessment

The RTTI approach demonstrates that each of the basic challenges for teachers that were mentioned in the introductory section, are met by means of a process that could be described as “scaffolded learning on the job” supported by an elaborately structured multi-media testing and teaching package. This is manifested by providing, background information, structures, cues and ranges of practical examples in the following areas:

a) Test construction
b) Diagnostic interpretation of test results
c) Support for providing feedback and didactic action planning
d) Implementation checks on newly initiated teaching and learning processes.

In addition, basic training on these facets of formative evaluation are provided as well, namely in the form of:

- E-learning and the Handbook RTTI for expertise about RTTI & OPSA, curriculum alignment, basic concepts of educational testing with practical check lists for formative evaluation and interventions in the classroom
- The webapplication RTTI-online which support teachers periodically for continuous improvement in their formative evaluation with important elements about monitoring curriculum alignment, the quality of test questions, didactic and pedagogical interventions.

References

Anderson, L.W. and Krathwohl, D.R. (2001) A taxonomy for learning, teaching and assessing. A revision of Bloom’s taxonomy of educational objectives. -hardcover Complete Edition - New York: Addison Wesley Longman, inc.

Andrade, Heidi. (2013). Classroom assessment in the context of learning theory and research. SAGE Handbook of Research on Classroom Assessment. 17-34. 10.4135/9781452218649.n2.

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. Review of Educational Research, 61, 213-223. doi:10.3102 /00346543061002213

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5, 7-74. doi:10.1080/0969595980050102

Block, A. de (1975) Taxonomie van leerdoelen. Antwerpen: Scriptoria.

Bloom, B.S. (ed.)(1956) Taxonomy of educational objectives: The classification of educational goals. New York: David McKay

Burns, M., K. and Symington, T., (2002) A meta-analysis of prereferral intervention teams: Student and systemic outcomes. Journal of School Psychology, 40, 437 -447

Datnow, A, Park, V. & Kennedy-Lewis, B. (2012) High school teachers' use of data to inform instruction. Journal of Education for Students Placed at Risk, 17(4), 247-265.

Davies, A., Herbst, S. and Sherman, A. (2016) Assessment for Learning: A framework for educator's professional growth and evaluation cycles. Chapter 14 in Laveault, D. & Allall, ed. (2016) Assessment for Learning: Meeting the challenge of implementation. Switzerland: Springer Nature.

Drost, M. and Verra, P. (2108a) Handboek RTTI. Bodegraven: Uitgeverijplus

Drost, M. and Verra, P. (2108b) Slimmer Leren. Bodegraven: Uitgeverijplus

Eraut, M. (1994) Developing professional knowhow and competence. London: Falmer Press.

Faber, J.M. and Visscher, A.J.(2014) Digitale leerlingvolgsystemen: een review van de effecten op leerprestaties. (Digitial pupil monitoring systems: a review of effects on student achievement results). Enschede: University of Twente

Folmer, E., Koopmans-van Noorel, A. & Kuiper, W. red. (2017) Curriculumspiegel 2017. Enschede: SLO – nationaaal expertisecentrum leerplanontwikkeling.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systemic formative evaluation: A meta-analysis. Exceptional Children, 53(3), 199-208.

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. London, England: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77, 81-112. doi:10.3102/003465430298487

Hendriks, M.A., Scheerens, J. and Sleegers, P.J., C. (2014) Effects of evaluation and assessment on student achievement: A review and meta-analysis. Enschede: University of Twente

Kim, S-E. (2005). Effects of implementing performance assessments on student learning: meta-analysis using HLM. (Unpublished Doctoral dissertation). University Park, PA: The Pennsylvania State University.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin, 119, 254-284

Laveault, D. & Allall, ed. (2016) Assessment for Learning: Meeting the challenge of implementation. Switzerland: Springer Nature.

Lyons, N. (ed.) (1998) With portfolio in Hand; Validating the new teacher professionalism. New York and London: Teachers College Press.

Mintzberg, H. (1979). The structuring of organizations. Englewood Cliffs: Prentice Hall

Nevo, D. (1995). School-based evaluation: a dialogue for school improvement. Oxford: Pergamon

Onderwijsinspectie (2015) Informatiegebruik voor kwaliteitsverbetering. url: rijksoverheid.nl/documenten/ rapporten/2015/01/08/informatiegebruik-voor-kwaliteitsverbetering

Onstenk, J. (2004) De school en de lerende organisatie; Denkend aan scholen zie ik lerende organisaties traag door oneindig papierland gaan. MESO magazine jrg 24(138).

Oprins, E. and Andriessen, J. (2002) Leerprocessen bij kennismanagement. Uit: Kennisnmanagement – leerprocessen organiseren in werkomgevingen (Dankbaar, M (red), Oprins, E. (red), Andriessen, J., Hoek, F. van, and Tonnenam, J.), CINOP Centrum voor innovatie van opleidingen, ISBN 9050033660.

Sanders, P. (red) (2013) Toetsen op school. Arnhem: Cito

Scheerens, J. (2004) The evaluation culture. Studies In Educational Evaluation, 30 (2). pp. 105-124. ISSN 0191-491X .

Scheerens, J. (2016) Educational effectiveness and ineffectiveness. A critical review of the knowledge base. Dordrecht, Heidelberg, New York, London: Springer.

Schildkamp, K., Karbautzki, L., & Vanhoof, J. (2014). Exploring data use practices around Europe: Identifying enablers and barriers. Studies in Educational Evaluation, 42, 15-24.

Shute, V. J. (2008). Focus on formative feedback. Review of Educational Research, 78, 153-189. doi:10.3102/0034654307313795

Sousa, D. A. (2009, 2nd ed.) How the gifted brain learns. Tousand Oaks, California: A SAGE Company.

Visscher, A. J., & Ehren, M. (2011). De eenvoud en complexiteit van opbrengstgericht werken. [Enschede]: [Universiteit Twente, Vakgroep Onderwijsorganisatie en -management].

Wiliam, D. (2011). What is assessment for learning? Studies in Educational Evaluation, 37, 3-14.

doi:10.1016/j.stueduc.2011.03.001
William, D. and Leahy, S. (2018) Formatieve assessment; integreren in de praktijk (bewerkt door: Pijl, P., Adrian, R. and Bont, l. de) Rotterdam: Bazalt Educatieve uitgaven.