

Critically Reviewing the Keeping Learning on Track Programme

Jiao Lu

Department of ESL, Primary School of Beijing Royal Experimental School, Beijing, China.

How to cite this paper: Jiao Lu. (2024). Critically Reviewing the Keeping Learning on Track Programme. *The Educational Review, USA*, 8(4), 580-586.
DOI: 10.26855/er.2024.04.016

Received: March 27, 2024

Accepted: April 23, 2024

Published: May 20, 2024

Corresponding author: Jiao Lu, Department of ESL, Primary School of Beijing Royal Experimental School, Beijing, China.

Abstract

Researchers in the field of education are convinced of the advantages of formative assessment, which is considered an effective approach to enhance students' achievement in modern education. However, some experts still argue about the validity and reliability of it. The purpose of this essay is to analyze the rationale behind the Keeping Learning on Track program, its definition, content, and how teachers implement its logical model. This report will aim to assess the validity and reliability of the data, examine its impact on teachers' classroom practices, students' educational engagement, and final achievement, critically analyze it and provide suggestions for future improvement and formative assessment. The result and conclusion indicate that there is still no evidence supporting the positive impact of the KLT program on student achievement. Therefore, additional empirical evidence is needed to advance the program and formative assessment.

Keywords

KLT programme, formative assessment, students' achievement

1. Introduction

Researchers and practitioners across the board of education are convinced of the advantages of assessments. If the assessment can support learning, it is labeled as “formative assessment” or “Assessment for Learning” (AfL). It is thought to be an effective approach to improve students' achievements and engagements in contemporary education (e.g., Black & Wiliam, 1998a, 1998b; Nolen, 2011; Stiggins, 2002). As Morris et al. (2021, p. 3) mentioned that there is still no singular definition for the term formative assessment. Wiliam (2006, p. 285) categorises formative assessments into long, medium, and short cycles. The short cycle assessment is regarded as the AfL. Black and Wiliam (2009) propose that “practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of evidence that was elicited” (p. 9). AfL, synonymous with formative assessment, focuses on the quality of learning instead of its outcomes (Schildkampa, K. et al., 2020, p. 2). An assessment may improve learning only if it gives feedback according to the evidence that has been collected from the observations of students' engagement and class work to evaluate themselves and each other, by teachers and their students. The primary purpose of this is to refine their classroom instructions, which therefore allows the assessment to become formative if the evidence is applied for modifying teaching instructions to adhere to the diverse needs of individual students (Black, Harrison, Lee, Marshall, & Wiliam, 2002, p. i).

Even though an increasing amount of evidence has implied that formative assessment may improve student attainment (Black & Wiliam, 1998b, p. 1; Nyquist, 2003), Bennett (2011) still argues about the validity and reliability of available evidence regarding the effectiveness of formative assessment instructions on enhancing student academic achievement.

The reasons for the uncertainties will be discussed further in the essay. In addition, Dunn and Mulvenon (2019) reviewed that limited scientifically based empirical evidence exists to support that formative assessment directly contributes to positive educational outcomes. Therefore, research led by Wiliam and the Educational Testing Service (ETS) produced the Keeping Learning on Track (KLT) programme. The on-going multi-year professional development programme is planned to guide teachers to employ the strategies of formative assessment every minute and every day during the class (Thum, Tarasawa, Hegedus, Yun, & Bowe, 2015, p. 1). When a meaningful definition of formative assessment is discussed, the KLT programme forms a practical community that attains to meet the needs with a theory of action, a logical model, and a concrete case to illustrate formative assessment. However, the KLT programme is also flawed in many ways, particularly the lack of robust data for its effectiveness.

The purpose of this essay is to analyse the rationale and the content behind the programme, taking into consideration how the logical model of the programme is implemented in classrooms by teachers. This report will also aim to detect its validity and reliability, to discuss its impact on teachers' classroom practice and students' educational experience and achievement, and to critically analyse it to offer some suggestions for future improvement. The first section details the introduction. The second includes the introduction to the KLT, validity and reliability of KLT, and following is the discussion of KLT's effectiveness on learning and teaching. Finally, the conclusion will synthesise key points, critically analyse the programme according to previous evidence, suggesting any future improvements in formative assessment.

2. Keep Learning on Track programme

2.1 The rationale for implementing the programme

An outstanding characteristic of assessment research is the shift towards focusing on the connection between assessment and classroom teaching and learning or formative assessment. This is away from the restricted forms of test or summative assessment that are weakly connected to students' learning experiences. Unlike the definition defined by Black and Wiliam (2009), Bennett (2011) further elaborates on the differences between "summative" and "formative" assessment (p. 7). If an assessment primarily aims to routinely evaluate the learning and the secondary to advance learning, it may then be summative. If it is well-designed, primarily supports for how classroom instruction should be adjusted or refined, whilst providing suggestions to teachers about students' understanding and skills, the same assessment may be considered as formative (see Table 1).

Table 1. The connection between the purpose and type of assessment (Bennett, 2011, p. 8)

Type	Purpose	
	Assessment of Learning	Assessment for Learning
Summative	X	x
Formative	x	X

Notes: X = primary purpose; x = secondary purpose.

Ample evidence related to the practices of assessment were examined by Black and Wiliam (1998a). The report indicates conclusively that formative assessment promotes students' learning, and significant gains appear in achievement (Black & Wiliam, 1998a, p. 61).

However, researchers have argued and criticised the claims regarding the benefits of formative assessment on learning performance and achievement gains (Kluger & DeNisi, 1996; Yin et al., 2008). Also, Yan et al. (2021) found that teachers' intentions and implementations, especially education and training might be a pivot point for promoting the benefits of formative assessment in students.

During the process of formative assessment, the evidence is garnered by teachers through various methods to identify the emerging understanding or skills of students, thus teachers can facilitate their growth by adjusting their instructions (Heritage, Kim, Vendlinski, & Herman, 2009, p. 24). However, the question I propose is that do all the teachers have the competence to adapt classroom instructions to meet students' diverse learning needs? The problem is illustrated through the results of a generalisability study of measures of teacher knowledge in mathematics teaching (Heritage, Kim, Vendlinski, & Herman, 2009). The results illustrated that teachers are better at making reasonable inferences to evaluate their students' understanding levels from classroom evidence but deciding next instructional actions is challenging due to their poor interpretation of data concerning their students.

Supovitz and Klein (2003) highlighted the need for sustained support from experts and ongoing professional development to equip teachers with comprehensive data interpretation and analysing skills to build teacher capacity that can implement formative assessment accurately. A poor implementation of formative practices in classrooms may result in inconsistent effects. In this regard, a practical programme such as the KLT can act as a provocative example that allows the advancement of AfL, elaborated at the ETS in 2010.

2.2 The content of the KLT programme

It was designed and developed by Dylan Wiliam, an expert in formative assessment, alongside his co-workers at the ETS in 2009 and 2010. It is a constant multi-year teachers' training programme, aiming to promote educators' formative practices, and student learning, and aid students in cultivating internal standards for their work, whilst reflecting upon it, and taking ownership of their learning (Bennett, 2011, p. 9). Typically, it involves a fundamental theory and a practical instance to exemplify formative assessment. Black and Wiliam (1998a, 2009) theorise that the theory focuses primarily on one big idea with five key strategies. The big idea is that students and teachers can apply evidence to modify learning and teaching instructions, enabling them to meet immediate learning needs minute by minute and day by day (ETS 2010, as cited in Bennett, 2011, p. 8).

The five strategies were first promoted by Leahy et al. (2005) who proposed the strategies to be questioning, sharing learning expectations, self-assessment, peer-assessment, and feedback. Questioning refers to teachers eliciting or collecting evidence of students' daily performance via designing effective discussions, questions, and assignments. Sharing learning expectations allows teachers to share and explain their intentions and standards for success. Self-assessment involves teachers offering opportunities for students to continue learning about their own learning. And peer-assessment allows teachers to structure opportunities for one another to guide students to be instructional resources. Feedback enables teachers to offer feedback to students as a method of propelling students' learning forward, whilst designing a structure for students to put it into action (see Figure 1). Teachers can adapt instructions that adhere to individual students' immediate needs through questioning thus, students' engagement may then be prompted. Through the following three strategies, students can also instruct each other and take ownership of their own learning. Additionally, students can act upon feedback after the teachers' provision of their feedback. These strategies are then deployed to evoke information to establish where the learner is (via questioning, self-assessment, and peer-assessment), where he or she is heading (through knowing teachers' expectations, intentions, and standards), and how to get there (by feedback) (Wiliam, 2004b, p. 5). Teachers or students can apply some classroom instructional techniques, such as "three stars and a wish" and "traffic lights", which implement the five strategies in the programme. "Three stars and a wish" is a checklist that is implemented with peer assessment. This is employed so students can exchange completed tasks and then each of them is expected to indicate three stars/points/things he or she enjoyed or liked the classmate's work, with one part stating any suggestions to better the work. For the "traffic lights", every student obtains three cups in the colours red, yellow, and green colour respectively. Then, everyone is asked to use the different-colour cups (red colour means "I don't understand"; the yellow one indicates "I am unsure"; and the green one means "please proceed") to indicate his or her level of understanding. Furthermore, it is worth noting that these conclusive strategies have connections with cognitive-scientific theory, to the area that links learning to social interaction.

On the left side of the KLT programme (see Figure 1), the components are postulated to elicit a reform in the teachers' classroom instructions or strategies which is shown in the middle. This can then affect student behaviour which ultimately enhances academic attainment, reflected on the right side. These components mainly focus on teachers' professional development in formative assessment which can be regarded as an intervention. The components include ancillary materials, programme modules, facilitated training, meetings, workshops, and continuous support from ETS consultants. The workshop is mainly for school staff members who desire to establish teacher learning communities (TLCs) for themselves. This would enable them to meet regularly in small groups to support interchange, reflection, and encourage exploring and enhancing flexible classroom practices (Harrison, 2005; Lee & Wiliam, 2006). Moreover, the training materials entail 16 modules that develop into the TLCs' curriculum, textbooks employed by the TLCs, as well as a guidebook for the leaders of the TLCs (Bennett, 2011, p. 10). The KLT postulates that the degree to which teachers attain is evidence of students' learning (see Path 1 in Figure 1) which should be improved day-by-day, all else being equal, for the teachers who participated in the on-going professional development (Thum, Tarasawa, Hegedus, Yun, & Bowe, 2015, p. 7). Thus, such improvement could positively influence student engagement (see Path 7 in Figure 1) and bring subsequent improvements regarding academic achievements (see Path 13 in Figure 1). Yes, "if measures of the individual component of the model are available, this set of predictions allows researchers to evaluate the effectiveness of the intervention by monitoring the post-intervention status of each of these components" (Thum, Tarasawa, Hegedus, Yun, & Bowe, 2015, p. 7).

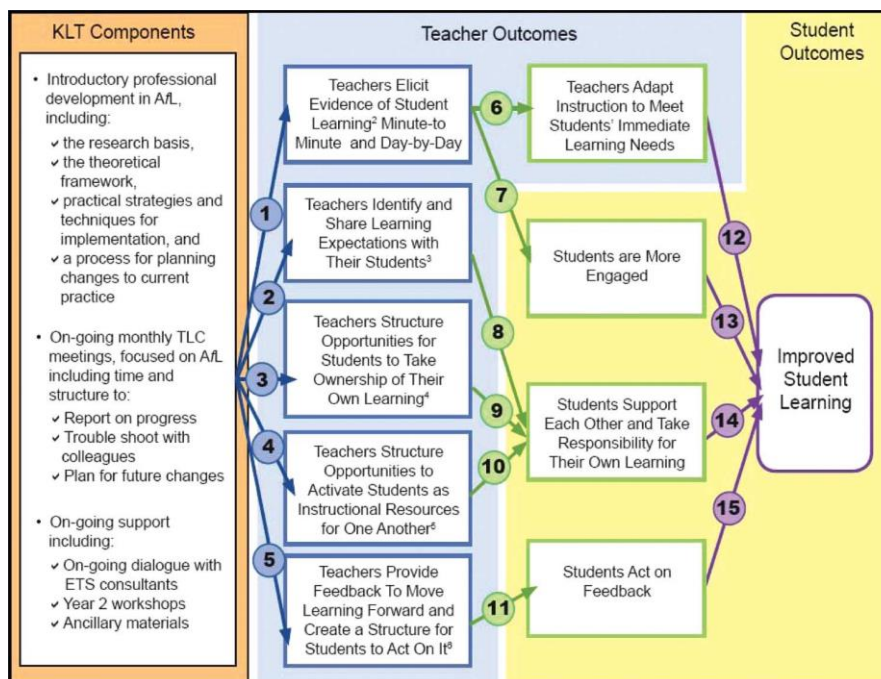


Figure 1. The KLT theory of action (ETS, 2009, as cited in Bennett, 2011, p. 9).

2.3 Validity and reliability of the KLT programme

The question I now propose is how can formative assessment be reliable and valid? When we talk about the summative purpose, the scores do not rely on the person who does the assessment – the assessment must be unbiased. Wiliam (2004a) stipulates that a key requirement for a valid and reliable formative assessment is that all teachers involved should have a shared construct of the quality of the curriculum for the subject in question. In this regard, the teachers getting involved in the assessment should develop a community of practice outside, and the formative assessment is to be implemented as intended to ensure the validity. Furthermore, Bennett (2011) claims that different observers made similar inferences about a student's understanding from the same evidence as the inferences drawn are consistent with other, more in-depth methods of characterising a student's knowledge and skills. And different observers modify the instructions similarly to the identical evidence (p. 14). These are in line with the professional development of the KLT programme which organises teachers across the school, district, and even the nation, to form a community of classroom practice. This therefore introduces the practical strategies for implementation and a process for planning changes.

Assessments can be reliable to the extent that they yield consistent outcomes, but the validity would connect with the inferences drawn about students' learning needs by the teacher, and the necessary actions to take. To achieve this, not only is a shared construct of quality needed, but an anatomy of quality is also quite necessary. For example, the teacher should not only assess a student's response but can break it down into components to promote improvement (Wiliam, 2004a). Additionally, Bennett (2011) suggests that making both formative inferences according to the evidence and subsequently adjusting instructions are significant, since a failure in either step may decrease the efficacy of AfL. If the formative inferences are wrong, the basis for modifying the instructions must be weakened. If the inferences are correct but the instructions are refined inappropriately, learning is also unlikely to happen (p. 14). In the KLT programme, teachers would report on the progress and troubles of students, and then plan for future changes in the ongoing monthly TLC meetings that focus on AfL. Henceforth, the teachers give feedback to feed learning forward and plan solid instruction for students in real-life classroom practices.

According to Wiliam (2004a), a focus on the role of formative assessment also throws the learner's role into clearer focus. To enhance effective learning, two further requirements are needed: 1) the assessments must aid the learners to fully understand the prerequisites and the standards of quality that the teachers have defined. In the KLT programme, teachers should identify and share their learning expectations with students. 2) The assessments must help students evaluate their own performance and their peers' performance, towards the goals that have been set out for them. Also, teachers provide students with chances to know about learning, as well as to activate students as instructional resources for one

another in the KLT logical model. In some formative assessment, teachers may only make conjectures based on their own observations. In the programme, the teacher-student interaction could make the validity of teachers' conjectures stronger, because teachers observe students' behaviour across multiple sources to refine their conjectures (Bennett, 2011, p. 17).

2.4 Effects of KLT on teaching and students' engagement and achievement

Black and Wiliam (1998a) assert that formative assessment promotes students' learning, and the gaining benefits in academic achievement is considerable (p. 61). However, available evidence regarding the influence of formative assessment in enhancing student achievement is still inconclusive. Thum et al. (2015) summarise the reasons for the uncertainties: first, few published studies were conducted in a way to help exclude competing explanations of intervention effects; second, some of the studies had small sample sizes which limited generalizability and lacked consistency in the process of the implementation of formative assessment; third, the diversity of achievement measures were applied in those studies (p. 4). According to Bennett (2011), to test the efficacy of formative assessment, data should indicate that formative assessment is implemented as intended. For instance, teachers who attend the meetings of TLCs and take time to share and critique their classroom formative practices in the KLT programme. Likewise, other intermediate outcomes are achieved, such as teachers presuming to share learning intentions to activate students as instructional resources for their peers in the model. Finally, students who participate in formative assessment change positively on outcomes of interest than those who do not (p. 15).

The effectiveness of the KLT programme was recently evaluated by a team of researchers from the Northwest Evaluation Association (NWEA) in 2015. Thum et al. undertook a survey-based case study to evaluate the impact of the KLT programme across 39 schools (20 for KLT, and 19 for the control group) at Meridian in the USA. The two-year study of the KLT involved teacher and student participants, and school participants in a control group. This was to compare the two groups with the programme elements and evaluate the effectiveness for student learning and achievement, as well as teacher practices. In the case study, both teachers and students were surveyed, and student achievement and growth data were examined during two academic years (from 2012 to 2014). Process evaluation was implemented throughout the study to ensure the fidelity of the KLT implementation. Positive results were shown from the study. The findings indicate that the KLT programme affected teachers' classroom instructions and student engagement over time. Moreover, the five strategies in the logic model were successfully applied according to the survey responses. The specific findings are as follows: first, a considerable difference in teachers' values and attitudes placed on formative assessment was found after the two years implementation of KLT. Teachers continuously elicited evidence of student learning and increasingly shared learning intentions or expectations and standards with students, which provided more opportunities for students to be familiar with their learning. Second, according to the survey responses from students' perspective, their learning engagement was increased alongside the teachers questioning, sharing learning expectations with them, and structuring opportunities to activate them in self-assessment and peer assessment. Students also pointed out that they understood what to learn, how to learn, and they were learning effectively and successfully during the study. In the intervention group, students displayed positive enhancement in learning over time, such as increasing involvement in discussions, and embracing and acting on feedback. Finally, there were no significant alterations in student academic achievement during the study. Overall, TLC leaders and teacher participants in the study demonstrated positive perceptions about their experience of implementing the programme. And sharing teaching instructions and useful experiences with other colleagues in the TLCs within the KLT programme was reported as the most helpful technique which supplemented their teaching.

3. Future improvement of KLT

3.1 The future improvement

The KLT programme offers a theory of action which is useful to implement in real-life classroom practice and to study for one category of formative assessment. Formative assessment should be an activity that is essentially to gauge how student learning is progressing while students are in the process of learning (Heritage, 2022, p. 9). The KLT programme predominantly adheres to the learning-involving aspect of practice. This method may be effective in practice; however, deep domain understanding is not incorporated into the programme, and measurement fundamentals are not formally addressed systematically.

In addition, time is another issue. Even if a practical method to guide teachers to build pedagogical skills, deep domain understanding, and a sense of the measurement fundamentals is found, time is nonetheless the most essential resource required for teachers to engage in the KLT programme (Bennett, 2011, p. 19; Wylie, 2008, p. 122). They need time to apply that knowledge, understanding, and skill into real practice. For instance, teachers may need regular time for their

TLC meetings, or time to share successful experiences and challenges and to receive peer feedback.

ETS collaborated with multiple school districts to implement KLT in the USA at the beginning of the 21st century. Researchers of the ETS staff members generalised the studies to describe the strengths and weaknesses of KLT (Leusner, Ellsworth, & Goe, 2008). First, stakeholders' intrinsic motivation emerged as a driving force for school reforms. Second, the positive attitudes towards KLT, particularly the provision of time and collaboration with other schools by the school principals, can promote its success of implementation. Moving on, the implementation of the KLT programme is deeply influenced by the local conditions without compromising elements of the programme, such as varying levels of teachers, administrative support, materials and resources, and assistance by ETS members. Fourth, the lack of time for training and teacher collaboration was an underlying theme in school districts' implementation of KLT, eroding the impact of the programme. Therefore, ETS determined the requirements for the programme implementation based on their studies in many school sites (Leusner, Ellsworth, & Goe, 2008). First, the core content included materials, workshops, and activities which must be the focus of the program and are non-negotiable. The sequence of workshops can be adapted to adhere local needs. Second, the programme participants' contact with teacher leadership is significant for teachers' collaboration and expertise cultivation. Third, the size and frequency of TLCs are critical as four to eight teachers in a TLC may promote a deeper knowledge of formative assessment. Teachers must meet monthly for at least two hours to maintain their focus on formative assessment. Administrative agendas should not intrude on proceedings. Finally, the teacher's choice is paramount. Teachers require the flexibility to determine which teaching techniques exactly they would like to incorporate into their practice.

3.2 Conclusion

To conclude, the KLT programme is based on the non-negotiable training elements and the built-in flexibility of the program which enables facilitators to adapt training sessions to the local needs. The meetings foster collaboration and encourage observation, reflection, and discussion of other's practice or experimentation with formative assessment teaching techniques (Wiliam, 2006). Nevertheless, there is still no evidence regarding the impact of formative assessment or the KLT programme on student achievement. Therefore, further empirical evidence is required to propel the programme and formative assessment forward.

References

- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa*, 1-13.
- Black, P., & Wiliam, D. (2009). Developing a theory of formative assessment. *Educational Assessment. Evaluation and Accountability*, 21(1), 5-31.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). Working inside the black box: assessment for learning in the classroom. London: King's College London Department of Education and Professional Studies.
- Dunn, K. E., & Mulvenon, S. W. (2019). A critical review of research on formative assessments: the limited scientific evidence of the impact of formative assessments in education. *Practical Assessment, Research, and Evaluation: Vol. 14, Article 7*.
- ETS (Educational Testing Service). 2010. About the KLT program. Princeton, NJ: Author.
<http://www.ets.org/Media/Campaign/12652/about.html> (accessed December 17, 2010).
- Harrison, C. (2005). Teachers developing assessment for learning: Mapping teacher change. *Teacher Development*, 9(2), 255-263.
- Heritage, H., Kim, J., Vendlinski, T., & Herman, J.L. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice*, 28(3), 24-31.
- Heritage, M. (2022). *Formative Assessment: Making it Happen in the Classroom*. Corwin Press.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback and interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership*, 63(3), 19-24.
- Lee, C., & Wiliam, D. (2006). Studying changes in the practice of two teachers developing assessment for learning. *Teacher Development*, 9(2), 265-283.

- Leusner, D. M., Ellsworth, J., & Goe, L. (2008). Scaling up across diverse contexts: Lessons learned from five implementations of the Keeping Learning on Track programme. Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education, Chicago, IL.
- Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, 9, e3292. <https://doi.org/10.1002/rev3.3292>.
- Nolen, S.B. (2011). The role of educational systems in the link between formative assessment and motivation. *Theory into Practice*, 50, 319-326.
- Nyquist, J. B. (2003). The benefits of reconstructing feedback as a larger system of formative assessment: A meta-analysis. Unpublished master's thesis, Vanderbilt University.
- Schildkampa, K., Fabienne M. van der Kleij, Maaïke C. H., Wilma B. K., Bernard P. V. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 1-16.
- Stiggins, R. J. (2002). Assessment Crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, 83(10), 758-765.
- Supovitz, J. A., & Klein, V. (2003). Mapping a course for improved student learning: how innovative schools systematically use student performance data to guide improvement. Consortium for policy research in education research report.
- Thum, Y. M., Tarasawa, B., Hegedus, A., Yun, X., & Bowe, B. (2015). Keeping learning on track: A case-study of formative assessment practice and its impact on learning in Meridian school district. Northwest Evaluation Association, 1-80.
- William, D. (2004a). Assessment and regulation of learning. Proceedings from the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- William, D. (2004b). Keeping learning on track: integrating assessment with instruction. Proceedings from the 30th annual conference of the International Association for Educational Assessment (IAEA), Philadelphia, PA.
- William, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11(3), 283-289. Retrieved from <http://www.informaworld.com/10.1207/s15326977ea1103&47>.
- William, D. (2008). Changing classroom practice. *Educational Leadership*, 67(4), 36-42.
- Wylie, E. C. (Eds.). (2008). Tight but loose: Scaling up teacher professional development in diverse contexts. ETS Research Report.
- Yan, Z., Li, Z.Q., Panadero, E., Yang, M., Yang L., & Lao, H. L. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in education: Principles, Policy & Practice: Volume 28*, 228-260.
- Yin Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M...Young, D. B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, 21(4), 335-359.