NWEA

Keeping Learning on Track

A Case-study of Formative Assessment Practice
and its Impact on Learning in Meridian School District

March 1, 2015

Y. M. Thum   B. Tarasawa
A. Hegedus   X. Yun   B. Bowe

Abstract

In partnership with Joint School District 2 in Meridian, Idaho, this theory-driven study assessed the impact of Keeping Learning on Track® (KLT™), a professional development program pioneered by Dylan Wiliam and his colleagues at the Educational Testing Service (ETS). A team of Northwest Evaluation Association™ (NWEA™) researchers surveyed teachers and students over two academic years (2012-13, 2013-14) to examine which elements of the program, and under what conditions, we saw the greatest impact on 1) the use of formative assessment instructional practices in the classroom and, 2) changes in student achievement measures in reading and mathematics. Our results were generally positive. The findings suggest that KLT influenced teacher practices and student engagement over time. Teacher and student survey responses report successful adoption of the five key formative assessment strategies. Additionally, participants reported positive evaluations of their KLT experience and cited sharing experiences with their peers in learning communities as the most helpful aspect of the program. There was little direct evidence that KLT impacted student achievement over the course of the study. But, as many researchers have observed, professional development programs may take more time to show discernable progress in terms of student achievement outcomes. Thus, it could be too early to see measurable gains in student achievement. NWEA plans to continue its partnership with Meridian and track student growth in subsequent years to explore this possibility. Lastly, various aspects of KLT implementation that were measured did not appear to influence our conclusions about KLT in Meridian. We hope our findings help Meridian and other partner districts assess the value of their investment in KLT and inform recommendations to improve the program and its delivery to improve student engagement and learning.

Executive Sponsor: Raymond Yeagley, VP Research and Chief Academic Officer
Principal Investigator: Yeow Meng Thum, Senior Research Fellow
KLT Research Team: Beth Tarasawa, Andy Hegedus, Yun Xiang, & Branin Bowe
Project Manager: Nate Young

Executive Summary

KLT, a professional development program pioneered by Dylan Wiliam and his colleagues at the Educational Testing Service® (ETS®), is a sustained multi-year professional development program which is designed to support teachers to adopt minute-to-minute and day-by-day formative assessment strategies. In partnership with Joint School District 2 in Meridian, Idaho, this theory-driven evaluation study assessed the impact of KLT on participating schools. Educators most often cite support of assessments if their use closely supports teaching and learning. Researchers and practitioners consider these types of assessments to be one of the more powerful ways to enhance student achievement and motivation available today.

However, much of the available evidence concerning the effectiveness of formative assessment instructional practices for improving student achievement remains inconclusive. One reason is that few of the published studies were designed in a way that ruled out competing explanations of intervention effects. Another uncertainty in judging the impact of formative assessment may be due to inconsistency in how formative assessment practices are implemented in these studies. Further complications arise from the wide variety of achievement measures employed.

This two-year study of KLT followed participant teachers and their students, as well as their counterparts in control schools, to compare experiences with KLT program elements and assess the impact on teacher practices and student achievement. Results will be as useful to stakeholders at Meridian as they are pertinent to the body of research on the pedagogical value of formative assessments.

We surveyed teachers and students over two academic years (2012-13, 2013-14) and examined student achievement and growth data from Fall 2011 through Spring 2014. At each phase of the study, we examined in which elements of the KLT program implemented in Meridian, and under what conditions, we find the greatest impact on 1) the use of formative assessment practices in the classrooms and, 2) changes in student achievement measures in reading and mathematics. The following six research questions are drawn from KLT's theory of action and guided our investigation in Meridian:

- What impact does KLT have on teacher formative assessment practices?
- What impact does KLT have on students' educational experience?
- What impact does KLT have on student achievement?
- How does KLT implementation vary in Meridian?
- Which components of KLT do teachers and Teacher Learning Community (TLC) Leaders value?
- How have teacher formative assessment practices influenced student educational experience and achievement?

Our results at Meridian were generally positive. The findings suggest that KLT influenced teacher practices and student engagement over time. Teacher and student survey responses report successful adoption of the five key formative assessment strategies. Specifically, the findings include:

- After completing two years in KLT, there was a noticeable difference in the value teachers placed on formative assessment.

- One year of KLT implementation was associated with a difference in teachers continuously eliciting evidence of student learning. This increase was maintained after year two as well.

- There was a significant increase in how often teachers shared learning expectations with students in the second year of implementation.

- One year of KLT implementation was associated with teachers providing more structured opportunities for students to take ownership of their own learning. This improvement continued in year two.

- One year of KLT implementation showed a difference in teachers adapting their instruction to meet students' learning needs, with continuing improvement in year two.

- As viewed from the student's perspective, the more often a teacher identified and shared learning expectations with their students and structured opportunities to activate students as instructional resources for one another, the more students reported experiencing increased engagement in their own learning.

Across the district, students reported that they knew what they were supposed to learn, knew how to learn, and were learning successfully throughout the duration of the study. In KLT classrooms, students reported high levels on engagement measures (e.g., students increased involvement in classroom discussions, welcomed and acted on feedback) with some positive increases over time.

The impact of these strategies on student achievement and growth was evaluated as well. We found Meridian students perform better than national norms for both achievement and growth before and throughout the duration of the study. In general, there were no significant changes during the course of the study in the overall averages for students of KLT participating teachers for which NWEA *Measures of Academic Progress*® (MAP®) data existed. Although an important goal for KLT is improved student learning, the immediate goals of KLT are to make formative assessment an integral part of classroom instruction. Therefore, near-term impact on student achievement for a district with already strong performance is not as likely as realizing improvements over the longer-term. We hope to continue our partnership with Meridian and track student growth in subsequent years to explore this possibility.

Overall, TLC Leaders and teacher respondents had positive perceptions about their KLT experience. Teachers and TLC Leaders respondents found the monthly meetings, personal action plans, sharing of experiences in the TLCs, and KLT modules used in the TLCs helpful. For both groups, sharing experiences with peers in the TLCs was the aspect of the program reported most frequently as very helpful.

Implementation fidelity, composed of a series of measures of how well the required or recommended elements of the intervention were assimilated in classrooms, varied across the two KLT cohorts. The first cohort significantly exceeded the KLT program expectations for time spent on TLC meeting each month. As a result they were able to complete approximately the same number of KLT modules in five months from the beginning of the KLT implementation to the end of the school year as the second cohort did in eight months.

Generally, we found that participants met their meeting and module completion requirements and recommendations. However, we also found that implementation of both recommended TLC activities were higher for participants who believed that formative use of assessment practices was a high priority for the district, school, and for themselves, and for those teachers who felt stronger administration engagement and support. These results suggest that critical program elements will tend to be better implemented when the district and

2

NWEA.

school administration are perceived to be clearly behind the intervention. Finally, we did not find that differences in implementation fidelity substantially influenced our study conclusions.

We hope our findings help Meridian, and other school districts, assess the value of their investment in KLT and inform recommendations to improve the program and its delivery to ultimately improve student engagement and learning.

Keeping Learning on Track:
A Case-study of Formative Assessment Practice and its Impact on Learning in Meridian School District

Y. M. Thum   B. Tarasawa   A. Hegedus   X. Yun   B. Bowe
Northwest Evaluation Association
Portland, OR

Introduction

In recent years, many educators have become roundly convinced of the benefits of assessments if their use closely supports teaching and learning. Researchers and practitioners consider assessments deployed towards this end, labeled as "assessment for learning" or simply "formative assessment," to be one of the more powerful ways to enhance student achievement and motivation available today (e.g., Black & Wiliam, 1998, 2010; Brookhart, Moss, & Long, 2008; Cauley & McMillan, 2010; Nicol & Macfarlane, 2006; Nolen, 2011; Ruston, 2005; Stiggins, 2002). However, much of the available evidence concerning the effectiveness of formative assessment (FA) instructional practices for improving student achievement remains in question (Bennett, 2011; Briggs & Ruiz-Primo, 2012; Dunn & Mulvenon, 2009; Kingston & Nash, 2011). One reason for the uncertainty is that, among the published studies, few are designed in a way to help rule out competing explanations of intervention effects. With the small number of studies available, part of the uncertainty in judging the impact of formative assessment may be due to a lack of consistency in how formative assessment is understood and its practices implemented in these studies. Further complications arise from the wide variety of achievement measures employed.

This two-year study of KLT, a professional development program designed to promote formative assessment practices in the classroom, followed participant teachers and their students, as well as their counterparts in control schools, to compare their exposure to and experiences of program elements to assess the impact on teacher practices and student achievement. Overall our findings suggest that KLT influenced teacher practices and student engagement over time. Teacher and student survey responses report successful adoption of the five key formative assessment strategies. Furthermore, teachers and Teacher Learning Community (TLC) Leaders had positive evaluations about their KLT experience and reported sharing experiences with their peers in learning communities as the most helpful aspect of the program. However, standardized assessment scores and growth measures provide little direct evidence that KLT impacted student achievement over the course of the study. Lastly, we measured various aspect of KLT implementation but, while they varied between teachers and derived support from the administrators, they did not appear to influence the results for KLT at Meridian.

The report is organized as follows. After introducing KLT within the broader research literature on formative assessments, we pose six key research questions, detail the study design and the sample realized, and identify and operationalize the key variables in KLT's theory of action. We then discuss the importance of measuring the fidelity of implementation in the context of our evaluation study. A detailed discussion of the results by research questions follows. We conclude with a discussion of the main findings, placed within the methodological challenges of the study, and suggest directions for future research.

Findings from this study inform recommendations to improve the KLT program and its delivery and extend our understanding of what conditions maximize chances of improved student engagement. Results also help partner districts assess the value of the investment in KLT and similar such professional development opportunities. Finally, it is important to recognize that research partnerships, such as the one with Meridian in this study, are positive examples of how NWEA supports its partners in a shared commitment to improve all student outcomes.

Background

What is Formative Assessment? Formative Assessment?

Until recently, a consensus on the very definition of formative assessments seemed elusive (Cech, as cited in Bennett, 2011). One way to better understand formative assessments is to resist a common but simplistic comparison with "summative" assessments, which are frequently seen as assessments that are specifically designed to evaluate performance. For one thing, according to Bennett (2011), an assessment may provide summative evidence if its primary purpose is to evaluate learning and secondarily to support learning (see Table 1). If appropriately designed, the same assessment may be regarded as formative if support for instruction is its primary purpose. What appears to be a more productive view is that an assessment may be regarded as formative or summative depending on what goal we attach to its use.

Table 1
A more nuanced view of the relationship between assessment purpose and assessment type.

Type	Purpose	
	Assessment *Of* Learning (AoL)	Assessment *For* Learning (AfL)
Summative	X	x
Formative	X	X

Note: X = primary purpose; x = secondary purpose. Reprinted from Bennett, R.E. (2011). Formative Assessment: A Critical Review. *Assessment in Education: Principles, Policy, & Practice, 18,* p. 8.

Researchers and practitioners now understand that formative assessment refers to the integration of assessment *with* instruction. For example, Popham (2008), like Black & Wiliam (1998, 2003, 2009), suggests that formative assessments refer to classroom assessment practices that inform learning and are thus more of a process, as opposed to being merely another configuration of test items. Researchers also need clearer guidance for identifying classroom practices that are formative because different, or poorly differentiated, implementation leads to inconsistent effects (Dunn & Mulvenon, 2009). To make real progress, Bennett (2011, p. 7) reasons that "well-designed and implemented formative assessment should be able to suggest how instruction should be modified, as well as suggest impressionistically to the teacher what students know and can do." A concrete step forward that advanced the cause of formative assessment emerged at the Educational Testing Service (ETS) in 2009.

Keeping Learning on Track®

Developed by leading formative assessment expert Dylan Wiliam and his colleagues at ETS, *Keeping Learning on Track* (Educational Testing Service [ETS], 2009, 2010) is a sustained multi-year professional development program designed to support teachers' adoption of five key formative assessment strategies. These strategies include: 1) teachers elicit evidence of student learning minute-to-minute and day-by-day, 2) teachers identify and share learning expectations with their students, 3) teachers structure opportunities for students to take ownership of their own learning, 4) teachers structure opportunities to activate students as instructional resources for one another, 5) teachers provide feedback to move learning forward and create a structure for students to act on it. Through a series of workshops and job-embedded engagement in school-based TLCs, KLT exposes teachers to a wide range of classroom techniques. At the core of the KLT professional development program is a definition of formative assessments as any curriculum-embedded assessment used by teachers to provide specific feedback to students about their progress and needs. More specifically, it is argued that *classroom practice 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 the evidence that was elicited" (Black & Wiliam, 1998, p. 7).

This conceptualization of formative classroom practice, the "big idea" behind KLT as Bennett (2011) calls it, is crucial because it offers an explicit theory of action which identifies what counts as a formative classroom practice, suggests how practice may be measured, and points to places we need to look at for its impact. That is, with KLT, we have a useful instance of a theory about formative assessments that is sufficiently well-specified to be implemented and its hypotheses about outcomes tested.

Figure 1 depicts the KLT logic model and outlines how the KLT components (workshop, program modules, and support materials) are designed to impact classroom and student outcomes (ETS, 2009). It considers professional development an intervention, describing the process by which collecting evidence of student understanding allows teachers to plan, adapt, and evaluate their instruction to address their students' learning needs.

KLT's theory of action predicts that teachers who receive sustained professional development make changes in their classroom practices (Butler, Novak, Jarvis-Selinger, & Beckingham, 2004). Specifically, successful professional development effects may be recognized by the increased use of five classroom strategies: teachers who elicit evidence of student learning on a daily basis can adapt their instruction (Jones & Krouse, 1988; Peterson, Carpenter, & Fennema, 1989; Fuchs, Fuchs, Hamlett, & Stecker, 1991), identify and share learning expectations, structure opportunities for students to take ownership of their own learning, activate students as instructional resources for one another, and provide feedback to move learning forward. These teaching practices should then influence student engagement (Jones & Krouse, 1988) and foster students' self-regulation of their own learning (Fernandes & Fontana, 1996). Finally, as student engagement increases, student learning improves (Finn & Rock, 1997; Skinner, Wellborn, & Connell, 1990).
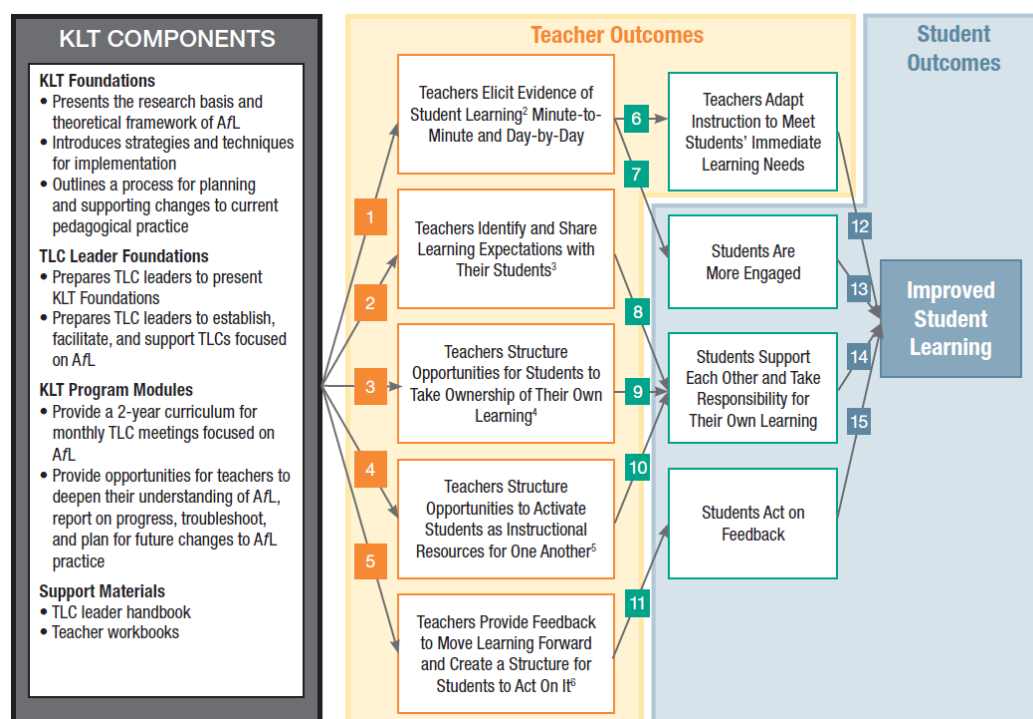
Figure 1. KLT logic model (ETS, 2009).

The logic model in Figure 1 represents the primary effects KLT anticipates for teachers and students. To take one example, KLT's theory of action predicts that the degree to which teachers elicit evidence of student learning (Path 1) should improve over time for those who participated in KLT professional development, all else being equal. Furthermore, such improvement should positively affect the level of student engagement (Path 7), and leads to subsequent improvements in students' academic outcomes (Path 13). What is important for understanding the impact of KLT is that, 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.

Current Evidence for Formative Assessment

Formative assessment draws from an array of compelling arguments regarding what makes a difference in teacher practice and student learning (Black & Wiliam, 1998; Brookhart, 2007; Guskey, 2007; Jones & Krouse, 1988; Peterson, et al., 1989; Fuchs et al., 1991). It is instructive to first understand the efficacy of the more popular forms of professional development before examining the effectiveness of formative assessment on student learning.

The research on the effects of professional development is varied. Some studies have shown professional development programs that include in-depth learning opportunities for teachers, content focus, and collective participation can be successful (Garet, Porter, Desimone, Birman, & Yoon, 2001). Other researchers have found that traditional forms of professional development are ineffective in providing teachers with sufficient time, content, or activities necessary for increasing teachers' knowledge or promoting meaningful changes in their classroom practices (Loucks-Horsley, Hewson, Love, & Stiles, 1998). Several studies have examined the

NWEA.

importance of specific dimensions of professional development, including how the duration of training impacts the depth of change in teacher practice and student achievement (Garet et al., 2001; Weiss, Montgomery, Ridgeway, & Bond, 1998).

Additionally, there has been growing interest in alternative forms of professional development such as mentoring, coaching, and professional learning communities (PLCs). Some researchers have argued that such "reform" types of professional development are more responsive to how teachers learn (Ball, 1996) and have a greater influence on changing teacher behaviors (Darling-Hammond, 1996; Sparks & Loucks-Horsley, 1989). The PLC model in particular represents a shift in the approach to professional development where PLCs are grounded in generating *knowledge of practice* (Cochran-Smith & Lytle, 1999). That is, PLCs employ both the knowledge and experience of teachers and the theory generated by researchers. Through collaborative exploration, PLCs support teachers in making decisions based on current and new professional knowledge, help setting goals, and ultimately to meet the needs of their students (Vescio, Ross, & Adams, 2008).

The past decade has seen increasing utilization of PLCs as integral both for school improvement efforts and as an important way to deliver effective professional development (Stoll, Bolam, McMahon, Wallace, & Thomas, 2006). Within schools, the phrase PLC is often applied to a group of teachers who are studying the same book, a school data committee, a group doing lesson study, or even a critical friends group. Recent literature suggests PLCs with certain attributes are effective tools for professional development. While not all grade-level teaming efforts result in better student outcomes (e.g., Saxe, Gearhart, & Nasir, 2001), many studies demonstrate that teams which are offered structured support for focusing on improving student learning can be highly effective (Saunders, Goldenberg, & Gallimore, 2009; Saxe et al., 2001; Vescio et al., 2006). Furthermore, Supovitz and Christman (2002) find that measurable improvement in student achievement only occurs in PLCs that are designed to change instructional practices of teachers. KLT utilizes TLCs, as the core site for the professional development intervention. These TLCs are teacher groups organized around KLT strategies and are thus functionally similar to PLCs.

If we turn to the research on the efficacy of formative assessment, we find that supporting empirical evidence remains inconclusive. Black & Wiliam's (1998) seminal work reviews more than 250 articles related to formative assessment. They claim that formative assessment improves learning and that student achievement gains are amongst the largest ever documented (effect-sizes of about .40 to .70). Other research corroborates (Brookhart, 2007; Guskey, 2007; Black & Wiliam, 2007; Hattie, 2008), suggesting that classroom assessment practices can have a sizeable impact on student achievement. Another effort, by Shavelson, Young, Ayala, Brandon, Furtak, Ruiz-Primo, Tomita, & Yin (2008), suggests that evidence from a large collaborative study on embedding formative assessments in inquiry-based science curriculum points to enhanced student outcomes.

However, Dunn and Mulvenon (2009) point to significant limitations in their finding. For example, some studies have included disproportionate number of students with disabilities (Fuch & Fuch, 1986), had small sample sizes (Martinez & Martinez, 1992; Whiting, Van Burgh, & Render, 1995), or included confounding assessment training (Bergan, Sladeczek, Schwarz, & Smith, 1991) which severely limit generalizability.

Furthermore, a recent meta-analysis by Kingston and Nash (2011) finds much more modest mean effect sizes (more likely to be about .20) and suggests that Black and Wiliam (1998) likely overestimated the impact on achievement. Briggs and Ruiz-Primo (2012) question several aspects to the Kingston and Nash (2011) study, but they agree that considerable uncertainty remains as to the impact of formative assessment on achievement. In

sum, while there is growing momentum behind the adoption of formative assessment practices in classrooms, research has shown that it will be important to measure the extent KLT is successfully delivered. Furthermore, the analyses by Bennett (2011), Briggs and Ruiz-Primo (2012), Dunn and Mulvenon (2009), and Kingston and Nash (2011) collectively suggest that the evidence-base for the impact of formative assessment on student academic achievement has remained weak. They specifically advocate two remedies: more appropriately designed studies, and research that examines the factors that influence the efficacy of formative assessment.

Meridian KLT Study Goals & Research Design

The principal goal of the study was to evaluate the delivery of KLT, and determine its impact on classroom practices and student learning. In selecting the research design for the study however, we first considered the characteristics of this type of intervention. We expected that changing classroom practices through professional development, such as KLT, would take time. Thus researchers should consider the anticipated time course, status of its delivery, and measureable impact in designing the study. Wiliam (2012) argues that impact through TLCs may be as many as 30 Program for International Student Assessment (PISA) points[1] between two or three years. Prior research suggests that program impact cannot be expected to be immediate or to unfold in a uniform manner across all participants, particularly when the study involves an intervention that is relatively extended in time (e.g., Fullan, 1997; Thum & Bhattacharya, 2001). For example, some participants may be quicker to experience an intervention fully because they may be more "ready" to take advantage of the new resources. It is also possible that logistical obstacles to implementation are lower or absent for them.

In light of these considerations, the guidance from KLT's theory of action was critical when we studied KLT in a naturalistic school setting. Overall, the project plan called for a two-year study, a period thought to be long enough for the intervention to take hold and for discernable results by the end of the study. We used 2011-2012 assessment data as baseline measures, began the intervention in the fall of 2012, and monitored its impact as it accrued into the spring of 2014. The study compared participating KLT teachers with a suitable control group of teachers on a number of post-intervention outcome variables from their baseline values. The study provided evidence with high ecological validity, employing a study design which satisfied the study inclusion criterion of Kingston and Nash (2011).

While it is important to address a question such as "Is KLT effective with respect to improving targeted outcomes, and if so, why?" the study was intended to offer more than a summative evaluation of KLT. The study also aimed to evaluate the delivery and success of the intervention in Meridian. That is, we were interested in answers to the question: "Has KLT been effective with respect to improving targeted outcomes *at this site*, and why?" Researchers and district administrators monitored how well major elements of the study were developing and, in some instances, adjusted treatment implementation (e.g., the addition of participants) with the shared goal of improving the chances of better outcomes for stakeholders. In short, the findings from this study not only contributed to our understanding of formative assessment practices but they give insights to our partners on whether and how KLT has impacted teaching practices and student outcomes at their site. For these reasons, our

---

[1] PISA scores have a mean of 500 and a standard deviation of 100. A 30-point increase amounts to roughly a "total" effect size of 0.33, or improvements that accrue from 0.10 to 0.15 per year.

study design is best thought of as an example of a formative evaluation (Scriven, 1967; Saettler, 1990). Finally, there was value for NWEA to conducting the study, as the results from this study are useful for product improvements. Furthermore, beyond contributing to the empirical evidence concerning formative assessment use and impact at Meridian, the study also developed evaluation skills and tools with which NWEA may better understand and deliver KLT in the future.

Research Questions

At each phase of the study, we examined which elements of the KLT program implemented in Meridian, and under what conditions, we find the greatest impact on 1) the use of formative teaching practices in the classrooms and, 2) changes in student achievement measures in reading and mathematics over two years (2012-13 and 2013-14). The following six research questions are drawn from KLT's theory of action (Figure 1) and guided our investigation in Meridian:

1. What impact does KLT have on teacher formative assessment practices?
2. What impact does KLT have on students' educational experience?
3. What impact does KLT have on student achievement?
4. How does KLT implementation vary in Meridian?
5. Which components of KLT do teachers and TLC Leaders value?
6. How have teacher formative assessment practices influenced student educational experience and achievement?

Study Variables

Treatment and Implementation

The participation status for a teacher or classroom is often thought of as the treatment or independent variable in the relationships examined. As we explained above, knowing how well individual teachers and their students embraced and utilized formative pedagogic practices holds the key to understanding KLT's impact on learning. Poor or incomplete implementation often leads the evaluator to mistake poor outcomes for an ineffective program (e.g., Hall & Loucks, 1978). Thus, we needed to know not just whether a teacher or a classroom was exposed to KLT, but also to what degree and whether exposure was accepted and received adequately. Implementation needed to be measured, monitored over the course of the study, and assessed for their impact on the evaluation as time-varying covariates (e.g., Fullan, 1983; Wang, Nojan, Strom & Walberg, 1984). We measured implementation quality by applying procedures such as the "fidelity index" (Cordray & Jacobs, 2005; Hullenman & Cordray, 2009) on information we collected about program participation and delivery, such as meeting attendance and adherence to specific program requirements.

Mediating Variables

KLT posits a role for mediating variables, which are classroom practices revolving around the use of feedback from suitable formative assessment techniques. KLT's theory of action (see Figure 1) outlines how embedded and sustained day-to-day teaching practices impact student learning; feedback via survey responses from students, teachers, and administrators were key data gathered for this study (please see Appendices B and C for survey

instruments). Each survey instrument operationalized a variety of core KLT concepts, including teacher perceptions, classroom techniques, and TLC engagement as conceptualized by the KLT program. We drew guidance from existing scales from other research domains (e.g., Babbie, 1990; Brace, 2004; Grunwald, 2012; Iarossi, 2006) and relied on a panel of experts to assist in developing items that measured each construct (McCoach, Gable, & Madura, 2013). Concepts included: knowledge of formative assessment, value of formative assessment practices, current instructional practices, professional development, organizational priorities, demographic background, implementation support, and student perceptions and practices. Surveys consisted of Likert-scale questions, multiple choice items, questions that allow respondents to choose one of many answer choices, and a small number of open-ended questions that allow respondents to answer in their own words. Survey item creation, question sequence, skip logic analytics, layout, and pretesting were conducted by NWEA researchers in collaboration with NWEA professional development consultants and district assessment leadership[2].

Table 2
Mediating Constructs and Outcomes in KLT's Logic Model.

Construct	Interpretation
S1SOAL	students support one another and take responsibility for their own learning
S2ENGG	students are engaged when they are involved in classroom discussions
S3WAFB	students welcome and act on feedback
T1EESL	teacher continuously elicits evidence of student learning
T2ISLE	teacher identifies and shares learning expectations with their students
T3SSTO	teacher structures opportunities for students to take ownership of their own learning
T4SSIR	teacher structures opportunities to activate students as instructional resources for one another
T5MLSA	teacher provides feedback to move learning forward and creates a structure for students to act on it
T6AISN	teacher adapts instruction to meet students' immediate learning needs
T7AISN	teacher professional development experiences and practices
CGI	Improvements from fall to spring in student MAP math and reading assessment results averaged over subject and grade-level

Table 2 lists the seven teacher and three student constructs central to KLT's logic model. KLT's professional development seeks to improve the degree to which teachers (1) elicit evidence of student learning (construct labeled as T1EESL), (2) identify and share learning expectations with students (T2ISLE), (3) structure opportunities for their students to take ownership of their own learning (T3SSTO), (4) structure opportunities to activate other students as instructional resources for one another (T4SSIR), (5) provide feedback to move learning forward and create a structure for their students to act on it (T5MLSA), and (6) adapt their instruction to meet the immediate

---

[2] A formal Institutional Review Board (IRB) at NWEA authorized the surveys and all partners in the study vetted every survey question.

learning needs of their students (T6PDVA). Additionally, we examined teachers' familiarity with the principles and strategies behind formative assessments, perhaps acquired through past professional development (T7AISN).

According to KLT's logic model, a general rise in the profile of measures over time likely reflects success in incorporating KLT into classroom practices. Furthermore, such classroom successes lead to improvements in how much students (1) support one another and take responsibility for their own learning (S1SOAL), (2) are engaged when they are involved in classroom discussions (S2ENGG), and (3) welcome and act on feedback (S3WAFB). These mediators are the effects of increased adoption of the formative use of assessments in classrooms. Therefore, they may be regarded as "leading indicators" of program impact as revealed, for example, by longer-term improvements in classroom average student learning gains within a school year. Specifically, the effectiveness claim of KLT professional development for teachers is that improvements in student outcomes are unlikely without meaningful improvements to these mediating factors in the classroom.

Outcomes

At various stages of the study, we measured KLT's impact on learning outcomes. We evaluated whether students in participating classrooms showed increasingly favorable achievement and growth patterns on NWEA's Measures of Academic Progress®[3] (MAP®) mathematics and reading tests when compared with students in control classrooms. The evaluation needed to take into account documented differences in the fidelity of implementation between classrooms as well as differences in simple proxies of classroom learning change among teachers as measured by the learning improvement collectively made by their students.

In this study, student learning improvement within each school year was indicated by the classroom average of the student *conditional growth index* (CGI). The CGI is a measure of how students grow in terms of NWEA's MAP assessments on mathematics and reading when compared with their peers, taking into account (1) how they individually performed at their initial assessment, (2) how they varied in terms of the amount of instruction they had received when they were tested, (3) their grade-level, and (4) the subject matter on which they are tested (Northwest Evaluation Association [NWEA], 2011). The CGI[4] is essentially a z-score[5]. Because the manner with which the CGI is standardized, we may compare two students on their CGIs regardless of their grade levels or subject matter. Its properties also make it suitable for arithmetic manipulations, such as aggregations. Each CGI has a normative position given by its corresponding conditional growth percentile, or CGP, for descriptive use.

Figure 2 represents the causal relations among the main classes of variables the study sought to understand. Generically, the original study design was to evaluate whether treatment (KLT components) impacted the mediating variables (classroom practices), which in turn shaped the outcome variables (student behaviors and learning), at various phases of the evaluation study. Next, we developed data collection and analytic approaches that were consistent with the design to address the above questions.

---

[3] MAP® is a series of computer adaptive assessments offered in the core academic subjects that are typically administered to students in grades two through ten. MAP scores, or RITs, have interval scale properties that support the measurement of academic growth.

[4] See Thum (in press) for more details on the CGI.

[5] A z-score is a standardized score that indicates its location in relation to the mean of the distribution from which it was drawn. Z-scores help researchers to compare scores from different distributions (Craighead & Nemeroff, 2004).

Figure 2. Causal links between treatment, its implementation, mediating, and outcome variables for KLT.

Sample

In this study, for each school participating in KLT, we employed the following general procedure to identify its "control" school[6]. First, we identified schools in the district, by the grade-levels served (primary, middle, high), their locale or urbanity (city, town, suburb, rural), their enrollment, and their School Challenge Index (SCI). The SCI is designed to provide a broader view of the "economic strain" schools experience as seen through a relevant set of socio-demographic, organizational, and educational policy programming factors (NWEA, 2011). It is used to compare schools in a state in terms of the collective economic circumstance of its students.

Next, among available candidate schools, a school which matched closest to the profile of the treatment school was selected for recruitment. The procedure resulted in 20 KLT and 19 control schools; a combined total of 39 schools. Matching on factors other than school locale, i.e., whether a school is in the city, in the suburb, or is rural, seemed appropriate. The unfavorable matching result for school locale is understandable given the size and school location realities of the district. Table 3 provides a count of the number of schools by key systemic or structural characteristics.

In year one, of those students with demographic information on gender and ethnicity[7], students in control schools were 52% males and 48% females. The proportion of males and females among KLT students are essentially identical. The student population was about 91% white or Asian[8]. Student MAP mathematics and reading scores for the fall of 2011 and the learning gains (CGP) they made through the spring of 2012 are given in Table 4. Student performance and growth appear to be quite comparable at baseline, although any differences were taken into account in subsequent analyses.

---

[6] Schools serve students with a variety of needs and operate in different contexts. As part of the selection process for control schools, and in an attempt to minimize the impact of these different contexts, the control schools were selected by matching them to participating schools.

[7] About 25% of students are missing demographic information.

[8] We group Asian Americans with white students because their academic achievement measures tend to be similar.

© 2015 Northwest Evaluation Association | 121 NW Everett St. Portland, OR 97209 | NWEA.org

Table 3
Characteristics of Participating and Control Schools Year One.

	School	KLT	Control
Location	City, Midsize	6*	3*
	Suburb, Large	11	6
	Rural, Fringe	3	10
School type	Regular School	18	18
	Alternative/Other School	2	1
School level	Primary	13	13
	Middle	3	3
	High	4	3
Title 1 eligible	Yes	7	6
	No	13	13
Magnet school	Yes	1	0
	No	19	19
Charter school	Yes	0	0
	No	20	19
FRL%	Average Percentage of Free-Reduced Priced Lunch students	32%	30%
School size	Average Total Number of Students	675	771

Note. *Without specific explanation the number in the two columns refers to school counts.

Table 4

MAP Mathematics and Reading Achievement and Growth for Sample Student at Baseline.

| | | Reading | | | | | Mathematics | | | | |
| | | Fall RIT | | | (%) | CGP (%) | Fall RIT | | | (%) | CGP (%) |
Group	Grade	n	M	SD	Median	Median	n	M	SD	Median	Median
	K	125	167	12.8	n.a.	n.a.	992	169	18.7	n.a.	n.a.
	1	1196	163	12.7	54	57	1195	165	14.9	59	51
	2	1225	182	14.0	68	43	1225	182	14.0	64	40
	3	1245	193	14.8	62	54	1246	192	11.7	53	78
	4	1139	204	14.0	65	48	1135	204	13.7	55	67
CTL	5	1164	209	14.4	62	58	1161	213	14.2	52	73
	6	1126	216	12.9	65	55	1128	222	12.6	57	61
	7	1048	222	11.9	67	58	1050	227	13.4	55	70
	8	942	226	12.9	69	53	955	233	14.9	58	69
	9	1115	227	12.6	65	56	1112	238	15.0	62	59
	10	7	234	12.8	79	15	1100	243	15.4	70	58
	K	136	168	13.8	n.a.	n.a.	792	170	18.8	n.a.	n.a.
	1	1079	162	12.8	53	57	1082	164	14.6	58	52
	2	1055	182	13.3	67	40	1055	182	13.5	64	36
	3	1073	193	15.1	62	57	1073	192	12.1	53	87
	4	1058	203	13.9	64	55	1060	204	13.0	54	74
KLT	5	1089	210	14.1	63	55	1085	213	14.0	55	76
	6	1178	215	14.5	61	56	1179	219	14.7	52	54
	7	1138	222	13.7	67	51	1138	228	15.6	56	62
	8	1136	225	13.9	67	50	1132	234	15.7	60	64
	9	888	227	14.1	67	56	894	237	16.9	58	63
	10	n.a.	n.a.	n.a.	n.a.		857	244	17.7	73	59

Note. n.a. denotes "not available."

In year two of the study, the district made the decision to roll out KLT in additional schools through its performance improvement plan which was a telling indicator that the district was committed to the program. This decision did complicate our study design and our ability to make definitive comparisons between participant and control schools over the two-year time period, a natural limitation of participant-centered evaluation research in schools. Thirteen of the control schools transitioned into participant schools (in year one of implementation) and additional KLT teams were formed in current participant schools. Thus, survey administration was limited to participants in year two. Table 5 presents the survey participants by subgroup for the duration of the study.

NWEA.

Table 5

Survey Participation by Subgroup (Count of Surveys Returned).

Surveys returned		Year/term			
		2012	2013	2013	2014
Who	Survey	Fall	Spring	Fall	Spring
Student	CTL	7,207	2,228	-	-
Student	KLT	8,033	3,331	2,297	1,923
Teacher/admin	CTL	165	89	-	-
Teacher/admin	KLT	330	142	195	211
KLT workshop	-	24	-	-	-

Method of Analysis

Three features of this study shaped our analytic approach. Of foremost importance in this study was time, because we sought to understand how key outcomes changed as KLT moved into the classroom over the course of the two-year study. A second feature of the study was the nested nature of the data, for which standard regression models would fail to represent the covariance structure of the data adequately and multilevel/hierarchical models (e.g., Raudenbush & Bryk, 2001) were designed to accommodate. A third feature was the adaptive nature of a formative evaluation.

This study's planned analyses would have evolved from models that provide single population descriptive statistics and simple tests of program effects to more complex models that seek to control for the impact of confounding variables. However, the unanticipated changes in the data discussed above had markedly increased the uncertainty in the information received that led necessarily to weakening the planned statistical treatment. Due to the mix of lower-than-ideal response rates, attrition, and also the policy-driven arrival of new enrollees to KLT, only rudimentary statistical comparisons such as exploratory bar-charts were attempted, to be supported by cross-sectional comparisons using t-tests with unequal variances or an analysis of covariance.

In the following sections, we outline the data and our approach to building measures of the constructs in KLT's logic model. Note that students take MAP reading and mathematics assessments during the fall and spring term each year. Their growth within the school year is important to assessing student learning. While gains in scores are typically hard to interpret by themselves, the CGI provided a sound basis for evaluating longitudinal performance of students and is available in the achievement and growth norms for NWEA's MAP.

Building KLT Constructs

Because the constructs in KLT's logic model were not directly observable, an effective measurement strategy was to triangulate using two or more observable indicators, and then validate its meaning by examining how the construct predicted other observed or unobserved variables. In many similar studies, responses to test or survey items provided the empirical referents to measuring and understanding the constructs proposed in a theory. In KLT's logic model, for example, it is hypothesized that the teacher construct T2ISLE, which represents the degree to which teachers identify and share learning expectations with students, is observable through the joint

responses to a combination of teacher (EC04f) and student survey items (SEC02a, SEC02b, and SEC02c) as depicted by the measurement model in Figure 3.

Tests of the measurement model for T2ISLE were accomplished, after initial exploration of the summary statistics and inter-correlations among the items, by using exploratory and confirmatory factor analyses with the items treated as indicator variables to measure the unobserved constructs (e.g., Child, 1990). The results either support or challenge the presumptions of a one-factor measurement model. Favorable results mean that the newly derived measure (named "T2ISLE"), a factor-analytic composite of the information contained the original survey item responses, can be used to represent the degree to which a teacher identifies and shares learning expectations with his/her students. We employed a similar strategy for the other constructs in the model.

Teacher Survey

ECO 4f. I identify and share the intended learning and success criteria with students.

Student Survey

SECO 2a. My teacher makes clear what is expected of me in a task.

SECO 2b. When I get something wrong, my teacher explains what I did wrong.

SECO 2c. My teacher makes sure I am clear about what we are going to learn and how we will know when we have learned it.

T2ISLE

How much a teacher identifies and shares learning expectations with his students.

Figure 3. Measurement model for construct T2ISLE.

Two important caveats about our decision to factor analyze the survey items are in order. First, factor analysis is a well-established approach to testing and refining constructs. Typically, factor analysis is well understood when data are sampled from a single population and less so when data have a clustered structure (i.e., when student survey responses are nested within teachers or classrooms). In the case of simple clustering involving the sampling of students each within a sample of teachers or classrooms, multi-level factor analysis can be employed (Reise, Ventura, Nuechterlein, & Kim, 2005). The challenges we encountered in this application are markedly different. For a construct such as T2ISLE above, a mixture of teacher and student survey items are deemed important to the definition of the construct. The authors are not aware of any strategy at this time for factor analyzing a matrix of responses from different levels of nesting. Therefore we pursue a pragmatic, albeit more simplistic, approach.

When only teacher survey items are involved, factor analysis was employed with the teacher-level survey response data. When only student survey items are involved in a construct, factor analysis was used with student-level response data and the resulting factor scores are aggregated in a weighted manner to the teacher-level. When responses from both surveys are involved, they are treated as two analyses, one for the teacher survey items and the other for student survey items. The factor scores for the student survey items are then aggregated

NWEA

to the teacher-level and represent the student-factor to be used with the teacher-factor derived from the factor analysis of the teacher survey items. Such constructs are to be represented jointly in subsequent analyses, if needed, by both their teacher and student components. When a construct contains only one item from either a student or a teacher survey, we presume that the item is equated with its construct and thus are not treated further with factor analysis.

The second caveat regarding our approach to factor analysis is that we have taken a "divide-and-conquer" approach. Our first task was to measure the constructs in the logic model. We measure each construct separately and then explore their relationships, rather than pursue a joint "measurement-plus-evaluation" typical of structural equations modeling, or SEM (Jöreskog & Sörbom, 1989; Bollen, 1989). The rationale for our approach was to render the evolution of relationships between constructs over time that are less contingent on the decisions we make about the measurement of the constructs themselves.

Measuring Student Achievement and Growth

As we have discussed above, a useful proxy for the classroom improvement in learning for a teacher is the average conditional growth index, or CGI, of his/her students. To reiterate, this metric provides a value-added measure that takes into account differences in initial student achievement as well as differences in the amounts of instruction students received. It is also especially useful in this evaluation where we are interested in comparing student learning gains in different subjects and in different grade levels. The CGI has been successfully employed with longitudinal MAP assessment data for both school and teacher evaluation (Thum & Xiang, 2010; Thum 2014, in press).

Baseline Measures

In this section, we present the results from the teacher and student surveys administered at the baseline, i.e., during the pre-intervention term in the fall of 2012. We also examine the observed relationships between the constructs and compare the results with relationships posited by KLT's logic model where participants had implemented the professional development regime successfully.

Statistical procedures such as chi-square tests were used to detect differences between KLT and control school teachers on their responses to items on the surveys. In terms of their means, T-tests were employed to compare KLT and control school teachers for teacher and classroom constructs. Lastly, given the complexity of the logic model and the moderate-sized baseline sample of teachers available, our goals are necessarily modest. We were satisfied with a more piecewise exploratory analysis to assess the profile of the potential outcomes of the intervention at baseline. At this stage of the investigation, we employed bivariate correlations and the analysis of covariance (ANCOVA) to study the relationships among constructs in the logic model, including their impact on classroom average learning growth.

Student Survey Items

Student surveys were conducted to collect 3rd - 8th grade[9] student perceptions and practices regarding their teachers' formative assessment practice (see Appendix C for the student survey instrument). The student survey consisted of 11 Likert-scale questions about students' view of various aspects of pedagogical classroom practice (using response categories "never", "sometimes," and "always") as well as their perspectives on whether they were learning and the way they learned (using categories "strongly disagree", "agree", and "strongly agree").

Over 15,000 students completed the student surveys at baseline. In general, students responded positively to all the questions raised. For example, the response "always" far outnumbered the response "never" and the response "strongly agree" far outnumbered "strongly disagree". This indicated that most students witness with some frequency classroom practices that might be deemed to be formative in their classrooms. A degree of social desirability implied in the manner and direction with which question content was stated, may have produced a skew in responses.

We also compared responses of both KLT and control teachers' students. We found that the two groups of students responded very similarly to most of the questions. Overall, the similarities between study and control school respondents at baseline reduced concerns about pre-existing differences in student and classroom dispositions.

Teacher Survey Items

We surveyed teachers to learn about formative assessment practices in their classrooms (see Appendix B for the teacher survey instrument). Between the two surveys, there were a total of 39 questions composed of many Likert-scale and other generic rating-scale items. Table A1 in Appendix A provides response frequencies as well as comparisons between KLT and control (CTL) teacher responses to the items. Note that there was considerable variability in response rates among the items. Among all the teacher survey participants, a number ranging from 156 to 315 responded depending on individual survey questions[10]. Respondents tended to employ the high end of the scale, due likely in part to the content and manner statements were worded. As with the student survey responses, the response "always" far outnumbered the response "never" and the response "strongly agree" far outnumbered the response "strongly disagree" in the teacher survey results. Overall, the pattern of responses appeared to suggest that most teachers hold positive attitudes toward formative assessment practices, and value the potential of formative assessment.

Surveyed teachers appeared to be informed about formative assessment. When asked what "formative assessment" evoked for them, teachers were less likely to identify it with giving test or quizzes (EC01b) and more likely to identify it with activities which include continuously gathering evidence of learning (EC01a), students doing guided reviews of their own work and that of their peers (EC01c, EC01e), and teachers providing students with feedback (EC01d).

---

[9] We limited student surveys to 3rd-8th graders for two primary reasons. First, we were concerned K-2nd graders could have literacy limitations. Second, many high schools do not administer MAP and thus, would limit our ability to examine trends in student achievement for this subpopulation.

[10] For our purposes in this report, only teachers who are matched with student survey information are included.

Surveyed teachers appeared to value and recognize the importance of key formative assessment practices. For example, the majority of teachers agreed, or strongly agreed, with the statement that formative assessment is valuable and important both for teaching (EC01c) and for student ownership and depth of learning (EC02a, EC02b, EC02d).

In terms of their own instruction, most survey teachers appeared to be frequently engaged in practices that promote the use of assessments in instruction (EC04a-EC04f). These practices included seeking feedback from students and involved them in class discussion about their understanding and effort for improvements. Roughly 80% of teachers said that they were always (category 5) or nearly always (category 4) engaged. We note that, for the above questions, there was little difference in response patterns between KLT and control school teachers. However, more control school teachers claimed that they more frequently provide instruction that is closely in tune with the learning needs of their students (EC04b).

However, and in contrast with their overwhelmingly positive responses to the above items, teachers were less likely to claim (about 50%) that that they always or nearly always provided the type of structure for students to reflect on, or used feedback to improve, their understanding (EC05a). Only 60% of teachers were just as positive that they structured opportunities for their students to act as instructional resources for one another (EC05b). The results for item EC06a, EC06b, and EC06c also suggested that it was less frequently the case for both KLT and their control school counterparts to intentionally incorporate into their instruction feedback and opportunities for reflecting on how, and how well, their students learned.

All teachers responded that they "agreed to strongly agreed" that they understood (EC07a), employed (EC07b), and received support for (EC07c) adopting formative assessment practices and strategies in their classrooms. Nevertheless, we found that control school teachers tended to rate more highly their understanding of formative assessment strategies or their involvement in bringing rigor to their plans for changing their practice (EC07a, EC07b). Teachers were also similarly positive about being accountable to peers (EC07d) and administrators (EC07e) for incorporating such practices in their teaching. Most teachers, generally 90% or above, considered it moderately to highly important for students to have opportunities to reflect on how (EC09a), how well (EC09b), or to choose how materials were learned (EC09c).

The level of positive endorsement to all the above questions painted the picture of a district that was knowledgeable about the principles and strategies of using assessment for learning. It was also reflected in how strongly teachers, from across the district, felt that formative assessments was a high priority to individual teachers (EC10c), to their school (EC10b), and to their district (EC10a).

Constructs in KLT's Logic Model

KLT's logic model was critical to framing the study in that it identified the "moving" parts of KLT and what outcomes the intervention seeks to change. Below, we report on our efforts to build measures of its constructs according to the item-construct "map" in Table 6, where an "x" maps an item to a construct. We hypothesized a number of survey items as the observed indicators for each of the unobserved KLT constructs given in Table A1 of Appendix A. Exploratory and confirmatory factor analyses were employed in tandem to test whether a one-factor measurement model adequately captured the covariation among the observed indicators (see Figure 3 above). Table A2 in Appendix A provides a summary of the factor analysis results.

Table 6

Item-Construct Map Showing the Locations of Survey Items in Hypothesized Constructs of KLT's Logic Model.

Survey	Item	S1SOAL	S2ENGG	S3WAFB	T0VAFA	T1EESL	T2ISLE	T3SSTO	T4SSIR	T5MLSA	T6AISN	T7PDVA
	2a*				x			x				
	2b				x							
	2c				x							
	2d				x							
	3a											
	3b											
	3c											
	3d											
	4a						x					
	4b*										x	
	4c							x			x	
Teacher	4d									x		
	4e								x			
	4f					x					x	
	5a									x		
	5b								x			
	5c											
	6a							x		x	x	
	6b							x			x	
	6c							x		x	x	
	7a											x
	7b											x
	7c											x
	7d											x
	7e											x
	2a			x			x					
	2b			x			x			x	x	
	2c			x			x					
	2d		x	x		x						
	3a		x	x					x			
Student	3b			x					x			
	3c		x	x					x			
	3d			x					x			
	4a	x										
	4b	x										
	4c	x										
	4d											

Note. * Low variance, not used. Color indicates items of a common factor. T4SSIR is measured by two factors.

On the whole, the results support the proposed mapping of item to the constructs in KLT's theory of action; there was one major factor in most of constructs (see construct map table and also factor analysis results Table A2 "Number of factors"). The major factors appeared to be sound composites of their respective items. The only exception was T4SSIR, which measured the degree to which a teacher structures opportunities to activate students as instructional resources for one another. There are two major factors, correlated at 0.234, in the construct. One related to teachers' perspectives on involving students in classroom discussion and helping them work with each other; the other concerned students' perspectives on classroom involvement and working with other students.

Assessing the Logic Model at Baseline

Table 7 provides summary descriptive statistics of the factor scores (as well as the items and teacher-level aggregates of the teacher base-year and first-year value-added measures Y1_CGI and Y2_CGI). It also offers test results of baseline differences between KLT teachers and their control counterparts on these measures. The profiles of mediating and outcome variables appear to be comparable for participant and control groups. This is

NWEA

welcome news for an evaluation that seeks to remove the possibility of alternative explanations due to pre-existing differences between the groups.

Table 7

Simple Descriptive Statistics for KLT Logic Model Items and Constructs for Respondents at Baseline.

Variable	n	M	SD	More** positive	Item/construct: a higher score on this construct more strongly indicates that
S1SOAL	197	0.081	0.240	CTL	students support one another and take responsibility for their own learning
S2ENGG	197	0.085	0.358	n.s.	students are engaged when they are involved in classroom discussions
S3FAWB	197	0.099	0.385	n.s.	students welcome and act on feedback
T0VAFA	311	-0.044	1.008	n.s.	teacher values formative assessment for improving teaching and learning
ec04f	295	3.463	0.740	n.s.	T1EESL, teacher continuously elicits evidence of student learning
SEC02d*	197	3.899	0.344		
T2ISLE	183	0.018	0.959	n.s.	teacher identifies and shares learning expectations with their students
T3SSTO	288	0.056	0.976	n.s.	teacher structures opportunities for students to take ownership of their own learning
T4SSIRa	134	0.188	0.723	n.s.	T4SSIR, teacher structures opportunities to activate students as instructional resources for one another
T4SSIRb	134	0.057	0.293		
T5MLSA	288	0.041	0.977	n.s.	teacher provides feedback to move learning forward and creates a structure for students to act on it
T6AISN	287	0.059	0.967	n.s.	teacher adapts instruction to meet students' immediate learning needs
T7PDVA	288	0.062	0.998	CTL	teacher professional development experiences and practices
Y1_CGI	574	0.220	0.545	KLT	Improvements from Fall 2011 to Spring 2012 student MAP math and reading assessment results averaged over subject and grade-level
Y2_CGI	613	0.250	0.613	n.s.	Improvements from Fall 2012 to Spring 2013 student MAP math and reading assessment results averaged over subject and grade-level

Note. * Classroom mean of student survey item. ** Either teachers or classrooms of KLT® or Control (CTL), or neither (n.s.), show more positive response overall on the constructs, based on t-tests for mean differences at the nominal $\alpha=0.05$ level. "n.s." denotes "not statistically significant."

It is helpful to consider the results for each construct before turning to the relationships between constructs. First, we examined the marginal distributions for the measures and considered differences between KLT and control school teachers. We noted some indications that (1) students in control school classrooms are higher in the degree to which they support each other and take responsibility for their own learning (S1SOAL), and (2) teachers in control schools have experienced greater exposure to professional development including perhaps formative assessments. Wherever appropriate, we adjusted for such baseline differences in later comparisons between KLT and their control school counterparts.

Of keen interest to this study was student achievement. We found that most students appear to be performing above their norms. The average CGI is above 0.0, and so we can conclude that students have generally gained above their 50% percentile growth rates. KLT classrooms appeared to be learning more as measured by the average CGI (Fall 2011-to-Spring 2012, regardless of grade level and subject) of their students. In subsequent analyses, we adjusted for such prior learning improvement differences in study sample classrooms. We also examined how constructs and outcomes were related at baseline (see Table A3).

Implementation Fidelity

It is widely recognized that faithful implementation can be difficult to achieve for interventions that, are, by their very nature extended in time and involve multiple actors who may play a role in deciding whether, or how fully, they participate in the program (Fullan, 1983). In this study, for example, some TLC Leaders or teachers may be absent for the required workshops, may not share their lesson plans, and/or provide feedback to each other as the program was designed. Under such circumstances, KLT's true impact cannot be properly determined if its required or recommended program elements are inappropriately delivered or are consumed by teachers in different ways. When evaluating the impact of KLT, therefore, we need to scrutinize elements of the realized program for signs of implementation infidelity.

Measuring Implementation Fidelity

One approach to enhancing the validity of the evaluation is to obtain a sound reading of the manner and degree with which KLT is implemented over the course of the study and weigh the influence of any differences in implementation on the evaluation of key outcomes (Wang et al., 1984; Cordoray & Jacobs, 2005; Century, Rudnick & Freeman, 2010). In this study, we quantified how teachers implemented the program and whether differences in implementation fidelity affected our measures of teacher classroom formative assessment practices, and student experiences of the use of assessment, and student achievement.

We created three measures of intervention implementation fidelity based on how participants are meeting different required or recommended benchmarks of the program. The KLT program specifies that high fidelity implementation has a number of minimum requirements, namely: 1) TLC Leaders are required to attend a three-day workshop (KLT Foundations, Presenting KLT Foundations, and Facilitating a TLC), 2) all teachers are required to attend a KLT Foundations workshop, 3) each month teachers and TLC Leaders attend TLC meeting(s) for a minimum of 90 minutes (this can be achieved in one or several TLC meetings as long as the total time meets the requirement), complete at least one module following the module sequence, and develop and implement personal action plans, and 4) in the summer between Study Year 1 and Study Year 2, participants are required to complete the Individual Study - Bridging Years 1 and 2. Additionally, it was strongly recommended that teacher participants share lesson plans, ask for and offer feedback to support others in embedding formative assessment in their teaching, participate in peer observations of formative assessment techniques, and co-lead TLC meetings. TLC Leaders and teacher survey items were then created and used to assess the extent to which the KLT program was delivered as intended.

To understand how implementation varied, the following implementation measures were created based on our logic model: 1) Teacher Learning Community (TLC) meetings and modules, 2) Required TLC activities, and 3) Recommended KLT activities. Each measure was simply the unweighted average of item responses. The exceptions were, in the case of TLC meetings and modules completion, the total meeting duration (the number of meeting multiplied by the estimated time of each meeting divided by recommended total duration) and the proportion of meetings attended to the recommended total. Table 8 shows the structure of the three components and the teacher survey items within them. We consider, in our discussion of the results below, how participants vary in their implementation of KLT and how these differences may have influenced our findings.

Table 8

Components, Sub-Components and Implementation Fidelity Index.

Components	Items	Minimal or required dosage/duration	Response options
Teacher Learning Community (TLC) meetings and modules	• Number of meetings completed • Duration of TLC meetings • Total Duration of TLC meetings per month • Number of KLT modules completed	90 min.* 5, 8 modules*	0-25+ 0, 30, 60, 90, 120, 150, 180, 210+ N/A- calculated from responses above
Required TLC activities	• TLC meeting attendance • Develop Personal Action Plans • Implement Action Plans • Feedback to support embedded FA**	5 5 5 5	1. Never 2. Seldom 3. Occasionally 4. Frequently 5. Always
Recommended TLC activities	• Share lesson plans • Watch peer(s) implement FA • Peer(s) watch me implement FA	5 5 5	1. Never 2. Once a Year 3. Once a Quarter 4. Once a Month 5. Once a Week
Other	• Co-lead the TLC meetings • Follow the module sequence		1. Never 2. Seldom 3. Occasionally 4. Frequently 5. Always

Note. * Recommended dosage/duration adjusted in 2012-13 for delayed implementation. ** The survey question "Feedback to support embed FA" uses the Response Options: "1. Never; 2. Once a Year; 3. Once a Quarter; 4. Once a Month; 5. Once a Week"

Study Results

This section presents the results from our analyses that examined the changes in use of formative teaching practices, student engagement levels, and student achievement measures in reading and mathematics. We organize and present our results by research question and we provide a series of comparisons that are helpful in examining differences in outcomes by participation subgroups.

Who were the study participation subgroups? A participation subgroup (PS) consisted of teachers who were involved with KLT with a particular pattern over the course of the study. The study collected four waves of information about a teacher's classroom(s): Fall 2012, Spring 2013, Fall 2013, and Spring 2014, which we designated as the Fall and Spring Terms of the 2012-13 and 2013-14 school years, respectively. In each wave, student MAP scores were associated with a teacher, along with the collective responses of her/his students' surveys, and were attached to her/his own responses to each teacher survey.

Table 9 outlines the three[11] participation subgroups that were compared across the research questions. Control group (CCCC) consisted of 221 unique teachers (31% of teachers), with a total of 866 (36% of records) teacher records of response data. These teachers are known to be in the control condition at each of the survey periods of the study. KLT Cohort 1 (KKKK) were KLT participant teachers at each of the survey periods of the study, and consisted of 323 teachers (46%) with 1263 records (52%). KLT Cohort 2 (CCKK) were control teachers in the 2012-13 academic year and became KLT participants in the 2013-14 academic year, and consisted of 159 teachers (23%) with 288 (12%) of the records[12].

Table 9
Realized Participation Subgroups.

Group	School year 2012-2013		School year 2013-2014	
	Fall 2012	Spring 2013	Fall 2013	Spring 2014
CCCC = Control group, teachers at all four time points (n=221)	CTL	CTL	CTL	CTL
KKKK = KLT Cohort 1, participant teachers at all four time points (n=323)	KLT	KLT	KLT	KLT
CCKK = KLT Cohort 2, control teachers in year one of the study who became participants in year two (n=159)	CTL	CTL	KLT	KLT

Note. KLT = participant group; CTL = control group

---

[11] Several additional subgroups emerged (e.g., participant in year one, not a participant in year two). However, three primary subgroups captured the majority of the study sample. All subgroups with fewer than five teachers were dropped from the analyses.

[12] The typical teacher in each of these subgroups data may be missing on several data elements, rendering the usual direct conventional analyses implausible.

Our analyses were designed to examine whether KLT was associated with improved outcomes across various participation subgroups over the two-year study. With an interpretable baseline, changes across PSs may be deduced by visual inspection aided by the "error bars." The error bars were 1.4 × Standard Error (SE) of mean estimates. Generally, means are not statistically different if their error bars overlap (Goldstein & Healy, 1995). Any increase in an outcome over the course of the study should be welcomed. However, increases in 2013-14 over levels seen for 2012-13 for KLT Cohort 1 was additional evidence in support of the hypothesis that KLT was having its intended impact. The evidence becomes more compelling when similar increases were seen for KLT Cohort 2, although lower subgroup numbers here mean greater caution is needed.

Research Question 1: What impact does KLT have on teacher formative assessment practices?

We first explore the impact of KLT on teacher-reported classroom attitudes and practices by examining teacher-level constructs. Here, we present our findings by construct and report differences across participation subgroups over time. At baseline, over 50% of the respondents reported they strongly agree (the top rating of the scale) that formative assessment was valuable: Construct T0VAFA, teacher values formative assessment for improving teaching and learning (teacher survey items 2b-d, see Appendix A). Furthermore, fall-to-spring increases were seen for both cohorts KLT Cohort 2 and KLT Cohort 1 from the first year of KLT participation. In KLT Cohort 1, we saw sustained levels into the second year. Figure 4 shows the rates remained high throughout the study for all participation subgroups but did increase between fall and spring each year. The impact of KLT on the value placed on formative assessment is mixed after one year of KLT. After completing two years in KLT, there was a noticeable difference in the value teachers placed on formative assessment.
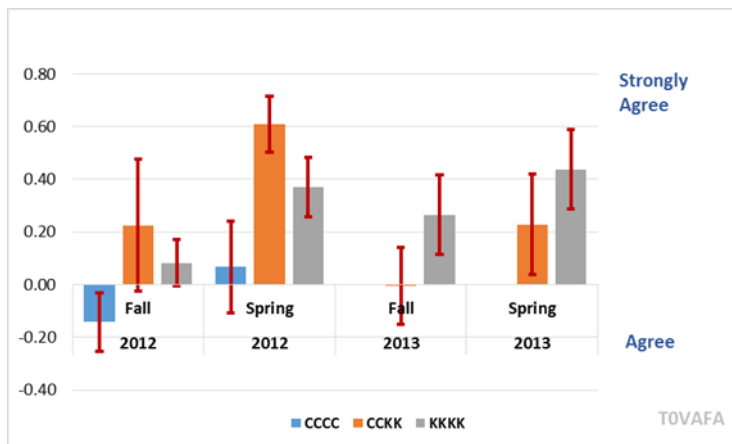
Note. 2012 refers to the 2012-2013 academic year, and 2013 refers to the 2013-14 academic year for all participation subgroup figures.

Figure 4. How strongly does the teacher value formative assessment for improving teacher and learning?

Construct T1EESL, "teacher continuously elicits evidence of student learning," combined one teacher survey item (4f) and one student survey item (2d)[13]. As can be seen from Figure 5, most teachers responded that they only occasionally or frequently collect responses from all students simultaneously. Year one of KLT implementation was associated with a difference in the solicitation of evidence. Both the KLT Cohort 1 and KLT Cohort 2 groups' differences from fall to spring in their first year of KLT implementation changed beyond the error band. This level was maintained after year two as well.
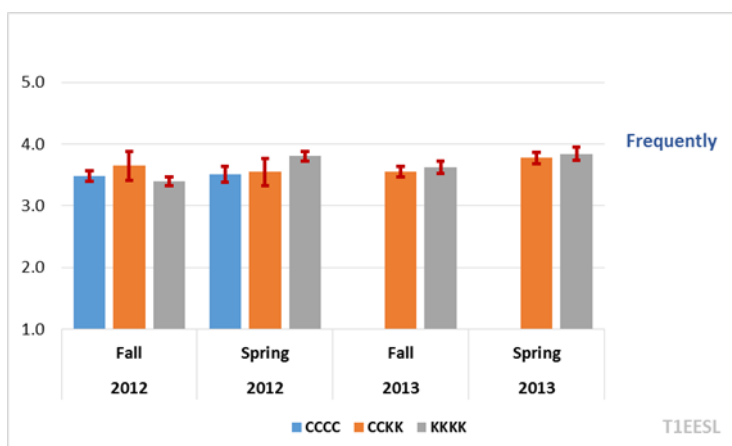
Figure 5. How often does the teacher continuously elicit evidence of student learning?

Construct T2ISLE, "teacher identifies and shares learning expectations with their students," combined one teacher survey item (4a) with three student survey items that overlap with Construct S2FAWB (2a-c). Thus, results were

---

[13] Generally, results hinged largely on the teacher survey item when student response rates were low. For the case of T1EESL, student information was unusually sparse.

similar between the constructs (see Figure 6). Interestingly, more students of Cohort 2 KLT classrooms (KLT Cohort 2) reported their teachers frequently ensure that they understand what high-quality class work looks like, their teacher frequently explains their errors, and were clear about their learning objectives and when they were successful. There was no change in year one of KLT with either cohort. However, in year two of KLT implementation, there was an improvement in how teachers shared learning expectations with students. Based on the sequence of the modules, with the learning expectation module given early in year two of implementation, these results were consistent with what would be expected.
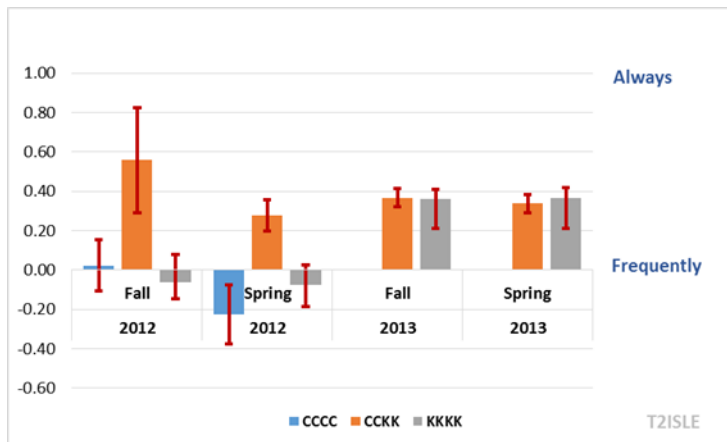
Figure 6. How often does the teacher identify and share learning expectations with his students?

Construct T3SSTO, "teacher structures opportunities for students to take ownership of their own learning," was composed of responses to teacher survey items 4c and 6a-c. Figure 7 shows the baseline for Control Cohort teachers appeared higher than the other two PSs, when teachers generally reported that they occasionally structured opportunities for students to take ownership of their learning. KLT was associated with teachers providing more structure opportunities in year one for both KLT cohorts. This improvement continued in year two. Over time, results suggest that KLT is associated with some marked improvement, with reports averaging about "frequently" by the spring of 2014 (about a 0.4 SD above baseline of 0).
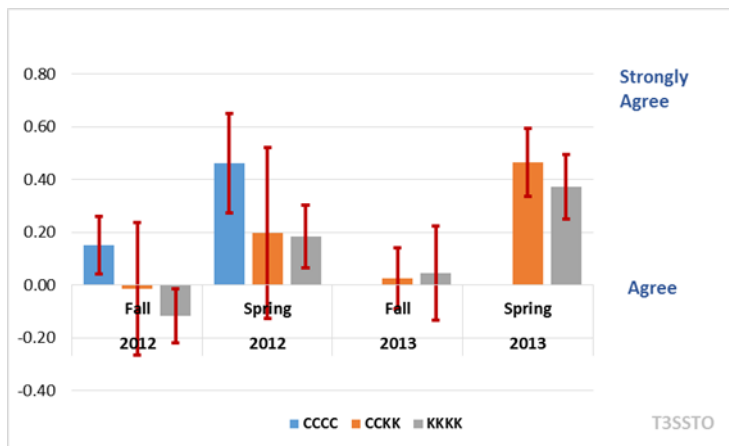
Figure 7. How often does the teacher structure opportunities for students to take ownership of their own learning?

Construct T4SSIR, "teacher structures opportunities to activate students as instructional resources for one another," was measured by two factors. The first was defined by two teacher survey items (teacher survey items 4e and 5b) and another by four items from the student survey (3a-d). This construct tapped how often opportunities existed in the classroom for students to serve as an instructional resource for each other. Figure 8 reveals at baseline, all teacher PSs reported that they frequently provided such opportunities. By the end of 2012 however, KLT teachers from both PSs reported that they frequently or always facilitated these opportunities. In contrast, Figure 9 shows students reported that such opportunities occurred only sometimes, but students from KLT classrooms reported much higher levels by the spring of 2014. Furthermore, the module that corresponds to this construct was completed in year two, thus, a meaningful change for the KLT Cohort 1 participation subgroup was consistent with KLT implementation. Although teachers reported a change in structured opportunities for peers to support each other, students did not report a significant difference in either year of KLT implementation. Please note that the vertical axes in the two figures below are not to be compared. Based on the changes within the KLT Cohort 2 and that of the KLT Cohort 1 PSs, the data suggests KLT may have helped increase such opportunities for self-directed and collaborative learning in the classroom.
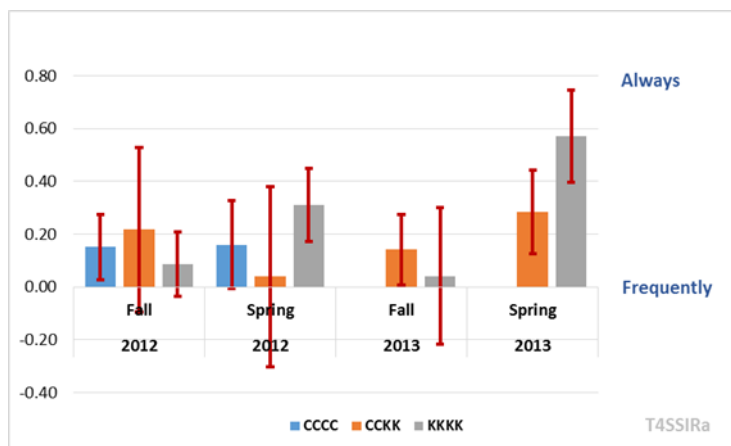
Figure 8. Teacher responses: How often does the teacher structure opportunities
to activate students as instructional resources for one another?
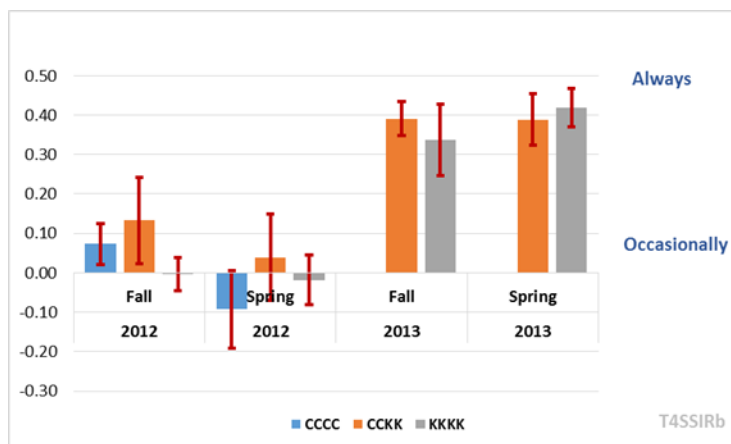
Figure 9. Student responses: How often does the teacher structure opportunities
to activate students as instructional resources for one another?

Construct T5MLSA, "teacher provides feedback to move learning forward and create a structure for students to act on it," was measured by four teacher survey items (4d, 5a, 6a, and 6c, please see Appendix A for more details) that provide a window into how the teacher reported effort in providing and helping their students use feedback. The construct also included one student survey item (2b), "how often does his or her teacher help with identified errors." Both teachers and students responded that they did so more than just occasionally[14]. Figure 10 shows overall, teachers reported that they occasionally provide and help students use their feedback on errors in 2012. Higher levels were seen for both KLT Cohort 2 and KLT Cohort 1 participant subgroups by the end of 2014 which suggests some support for the hypothesis that KLT impacted the teacher practices in question over time.

---

[14] These were not highly correlated.

NWEA.
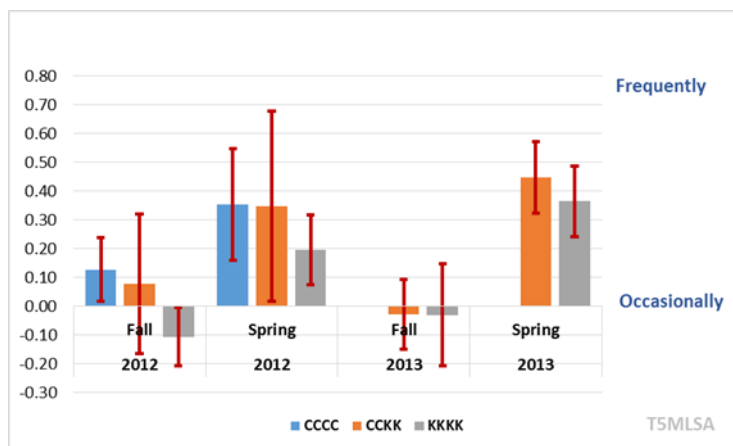
F*igure 10*. How often does the teacher provide feedback to move learning
forward and create a structure for students to act on it?
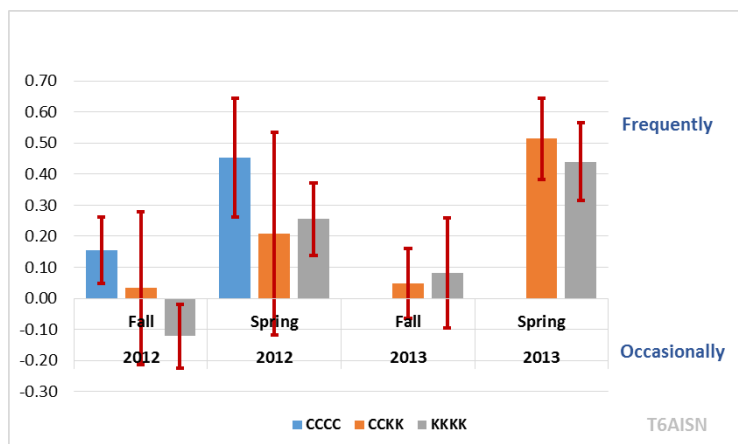
Figure 11. How often does the teacher adapt instruction to meet students'
immediate learning needs?

Construct T6AISN[15], "teacher adapts instruction to meet students' immediate learning needs," was measured by
six teacher survey items (4b, 4c, 4f, 6a-c) and one student survey item (2b). Figure 11 shows all teacher subgroups
reported adapting their instruction for students, with control teachers reporting the greatest frequency. However,
it was encouraging that the practices saw marked increases in frequency in the spring of 2014 in KLT classrooms.
Year one of KLT implementation showed a difference in teachers adapting their instruction to meet students'
learning needs for both cohorts, with continuing improvement in year two.

Construct T7PDVA, "teacher professional development experiences and practices," was measured with five
teacher survey items (7a-e). Figure 12 shows, overall, teachers agreed that they understood, took advantage of,
and received support for professional development in formative assessment. Additionally, they agreed that they
were accountable to their peers and to their administration to incorporate formative assessment strategies and
techniques into their classrooms. There was a noticeable increase in the level of agreement for Cohort 2 KLT

---

[15] Construct T6AISN shares some of the same survey items as prior constructs. However, KLT's theory of action suggests these
items combine to form an important construct that measures instructional adaption.

participants in the spring of 2014 and KLT classrooms showed an overall increase of about a half standard deviation over the baseline levels.
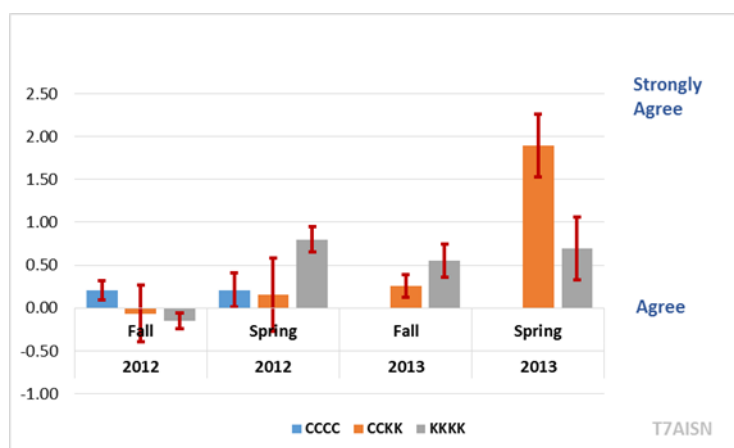
Figure 12. How strongly do teachers agree that they are engaged with, supported by, and benefitted from professional development practices?

Research Question 2: What impact does KLT have on students' educational experience?

We also examined the impact of KLT on student-reported classroom attitudes and practices. Overall, we found the educational experiences of students in KLT classrooms remained very positive when judged against baseline figures, or measurably improved as KLT practices took hold in these classrooms. We present our findings by construct and report differences across participation subgroups over time. Construct S1SOAL, "students support one another and take responsibility for their own learning" (student survey items 4a, 4b, and 4c), indicated that students strongly agree that they knew what they were supposed to learn, knew how to learn, and were learning successfully. Figure 13 suggests that there were no discernable improvement or differences across the participating subgroups, where PSs topped the available scale at baseline and stayed there over the course of the study.

32
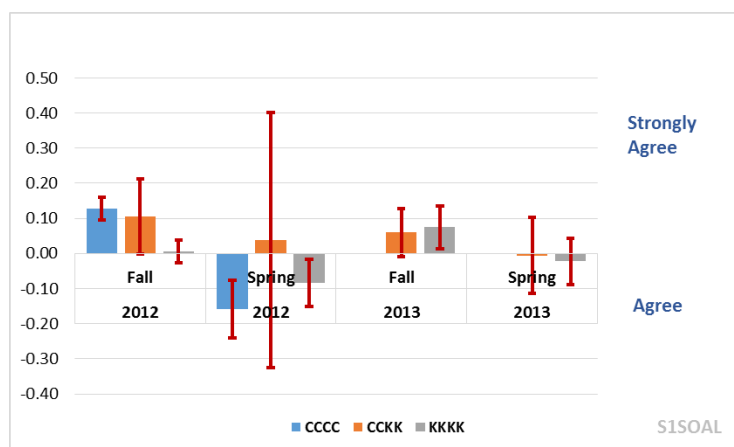
Figure 13. How strongly do students agree that they support one another and take responsibility for their own learning?

Students generally reported that their teachers frequently made sure that they respond to her/his questions in some manner (student survey item 2d), and that students are involved (student survey item 3a), and exchange suggestions to improve results (student survey item 3c). Overall, at baseline, all PSs were average on Construct S2ENGG, "students are engaged when they are involved in classroom discussions." Participant subgroups, KLT Cohort 2 and KLT Cohort 1, saw no discernible change from fall to spring in both years. (see Figure 14).
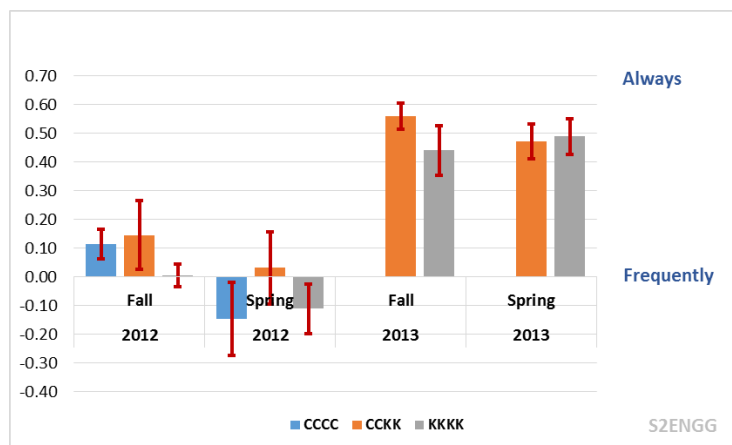
Figure 14. How often are students engaged during classroom discussions?

The third student-level Construct S3WAFB, "students welcome and act on feedback," consisted of six items (student survey items 2a-d and 3 b-d). Similar to the prior student engagement construct, overall levels for all three PSs were relatively high (measured in terms of frequency). Again, Figure 15 points to no discernible changes from fall to spring, but we did see an increase in student engagement from year to year.
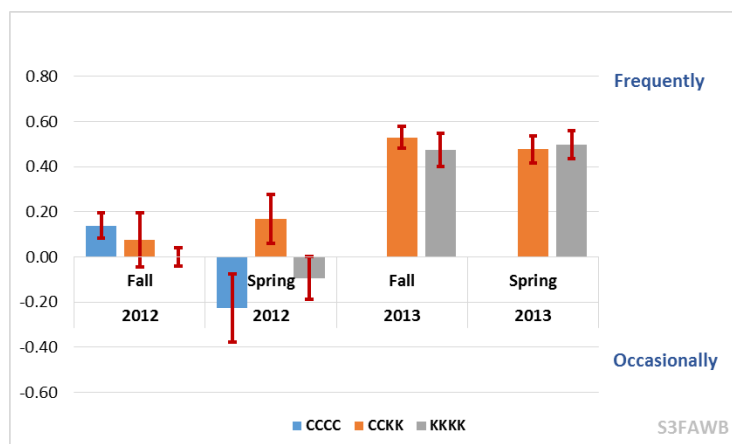
Figure 15. How often do students welcome and act on feedback?

Research Question 3: What impact does KLT have on student achievement?

Next, we examine student achievement patterns across all participation subgroups. Teachers from all groups tend to serve students achieving at about the 60th percentile (regardless of subject and grade-level). An exception is the lower achievement for classrooms of Cohort 2 study participants, whose students achieved at about the 50th percentile in the Fall of 2013. Nevertheless, student achievement of Cohort 2 KLT teachers inched upwards by the end of their first year of KLT in the spring of 2014. Figure 16 presents the achievement percentiles by participation subgroup over time. In terms of growth, as measured by the conditional growth index (CGI), achievement gains in the study classrooms were very stable over the course of the study, also at about the 60th percentile throughout (Figure 17). Overall, achievement and growth were above average for the district, with no noticeable changes in achievement or growth over time for any PSs. There was little direct evidence that KLT has impacted achievement by the spring of 2014.
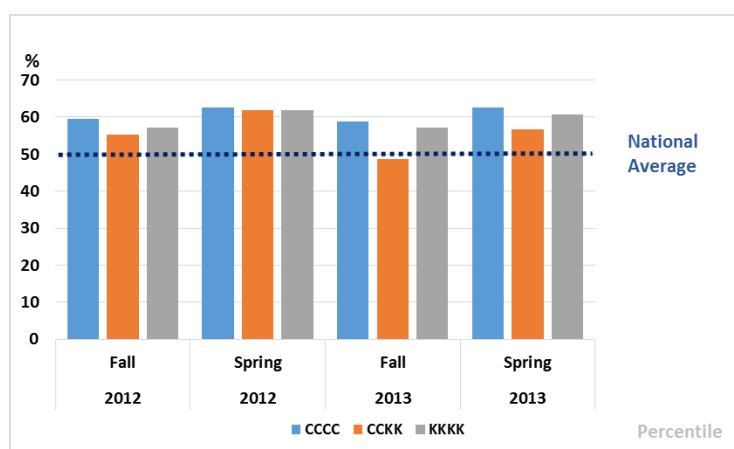
Figure 16. Average percentiles over grades and subjects.
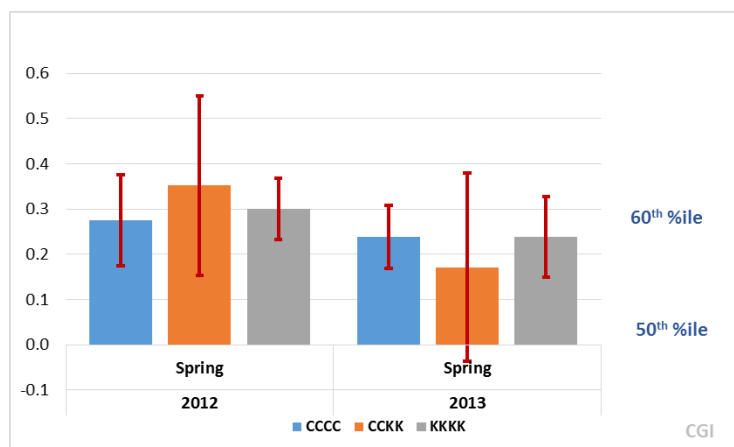
Figure 17. Average fall-spring gains over grade levels and subjects.

Research Question 4: How does KLT implementation vary in Meridian?

As discussed above, we measured implementation fidelity with three indices. We found that, on average, teachers and TLC Leaders of the two study years reported similar numbers of meetings and KLT modules completed although TLC meetings during the second study year were significantly shorter than those in the first study year. Because the KLT program in the district started late in the first study year, coordinators of the intervention set different benchmarks that would guarantee a minimum of 90 minutes per month on TLC modules. The benchmark number of modules are also changed; five modules for study Year 1 (2012-13 and KLT Cohort 1) and 8 modules for study Year 2 (2013-14 for both KLT Cohorts). The results indicate that TLC teachers and TLC Leaders completed the same number of modules in five months for study Year 1 and in eight months for study Year 2. Teachers and TLC Leaders noticeably overachieved and exceeded the benchmark set for them in the first study year, and both cohorts mostly met their benchmark second study year.

As for KLT-required TLC activities, teachers and TLC Leaders all responded very positively to the four questions. The responses by teachers and TLC Leaders in Spring 2014 were similar to those who filled out the survey in the previous year. For program recommended TLC activities in spring 2013, teachers shared lesson plans slightly less than "once a month" on average. However, the frequency dropped to less than once a year for questions: "watch a peer implement formative assessment plan" and "peers watch me implement plan." Compared with responses in Spring 2013, teachers and TLC Leaders in Spring 2014 reported slightly higher frequencies in "watch peer(s) implement FA" and "Peer(s) watch me implement FA".

We found that, on the whole, teacher and TLC Leaders more than met their combined meeting time and module completion recommendations. When they started, Cohort 1 participants were at twice the recommended amount and held an edge over Cohort 2 participants. However, it appeared that Cohort 1 teachers and TLC Leaders had spent less time in meetings and completing modules in their second year of implementation. For required TLC activities, the teachers and TLC Leaders had very similar responses, both indicated that they acted only frequently rather than always. In terms of recommended TLC activities, teachers and TLC Leaders from both cohorts reported that they share lesson plans, or watch each other implement formative assessment practices, less than weekly as recommended. Lastly, there appeared to be little changed in their pattern of participation of TLC activities from the first to the second implementation year, the only exception being Cohort 1 teachers and TLC Leaders appeared to have slightly increased their participation in recommended TLC activities. In sum, implementation appeared to be less than the ideal but cannot be judged to be inadequate, and to a large extent uniformly so for all KLT participants.

Not only were the averages largely similar, variability among teachers and TLC Leaders appeared to depend on the implementation fidelity measure in question. Participants varied between 16% and 30% on how much of their energies were expended in TLC total meeting duration and modules completed for the academic year. These percentages were the high and low coefficients of variation, or C.V., for the KLT cohorts in the study[16]. However, participants varied little when it came to required TLC activities; its standard deviation was well within one rating

---

[16] The C.V. is the ratio of the standard deviation of a sample of observations to its mean. It is frequently employed as a relative measure of variability in a sample. The lower the C.V., the lower the variability in a sample.

scale unit and its C.V. was only about 15% or thereabouts. Perhaps understandably so, participants varied about twice as much (C.V. of nearly 30%) when the TLC activities were recommended.

Additionally, we found the three measures were not highly correlated. In the first study year, the correlation between implementation fidelity measures for TLC total meeting durations and modules completed with required TLC activities was about 0.3, the correlation between implementation fidelity measures of TLC total meeting durations and modules completed and recommended TLC activities was about 0.13, and that between required and recommended TLC activities was 0.34. All correlations among these three measures were noticeably lower during the second year. This led us to suspect that, by the second year, the duration of the TLC meetings and the number of KLT modules completed would not be strongly related with teachers' involvement in KLT (such as TLC meeting attendance, developing and implementing personal action plans, providing feedback, sharing lesson plans, and observing peers). Thus it appeared that the three measures may provide non-redundant and perhaps useful insights into how well various aspects of KLT have been implemented in the district.

Table 10
Strong Administrative Support for Formative Assessment and Administrative Engagement Increased KLT Implementation.

	Recommended meetings & modules	Required TLC activities	Recommended TLC activities
FA high priority for district	n.s.[1]	0.26[2]	0.30
FA high priority for school	n.s.	0.46	0.37
FA high priority for teacher	n.s.	0.34	0.19
Administration engaged/support	n.s.	0.25	0.41

Note. n.s. [1] denotes not statistically significant; [2] correlation.

What might have influenced implementation of KLT program elements in the study? We examined several survey items that tapped participants' assessment of district support for KLT. From the correlations in Table 10, we found that participants uniformly met their meeting and module completion recommendations. However, we found that implementation of both recommended TLC activities were higher for participants who believed that formative use of assessment practices was a high priority for the district, school, and for themselves, and for those teachers who reported feeling stronger administration engagement and support. These results suggest that critical program elements will tend to be better implemented when the district and school administration are perceived to be clearly behind the intervention.

Research Question 5: Which components of KLT do teachers and TLC Leaders value?

A fuller appreciation for a professional development intervention such as KLT must identify the components of KLT that teachers and TLC Leaders value. As a part of the study, we included a variety of items in our final survey administered in Spring 2014 to assess the perceptions of TLC Leaders and teachers of various elements of the KLT program. Table 11 presents the descriptive information. We also include Net Promoter® scores to determine whether teachers and TLC Leaders participants were: "Promoters"- participants/customers who are highly likely

to recommend a product, "Passives"- satisfied but unenthusiastic participant/customers, or "Detractors"- people who are less likely to recommend a product.

Overall, TLC Leaders and teacher respondents had positive perceptions about their KLT experience, with more Leaders reporting KLT components were very helpful. All TLC Leader respondents found the monthly TLC meetings, personal action plans, sharing of experiences in the TLCs, and KLT modules used in the TLCs helpful. Teacher survey respondents also found the same four aspects helpful. For both groups, sharing experiences with peers in the TLCs was the aspect of the program reported most frequently as very helpful.

Table 11

KLT TLC Leaders and Teacher Perceptions about Aspects of KLT.

Variable	Respondent	n	Very helpful	Helpful	Not helpful at all	N/A	% Positive	Net promoter
Foundations day	Teacher L.	36	61%	22%	3%	14%	83%	58%
	Teacher	105	23%	28%	9%	41%	50%	14%
Monthly KLT TLC meetings	Teacher L.	37	81%	19%	0%	0%	100%	81%
	Teacher	102	74%	19%	3%	5%	92%	71%
Developing Personal Action plans	Teacher L.	26	58%	42%	0%	0%	100%	58%
	Teacher	76	54%	34%	5%	7%	88%	49%
Sharing experiences with peers in TLCs	Teacher L.	36	89%	11%	0%	0%	100%	89%
	Teacher	102	75%	16%	1%	9%	90%	74%
KLT modules used in TLCs	Teacher L.	27	67%	33%	0%	0%	100%	67%
	Teacher	81	57%	33%	2%	7%	90%	54%
KLT videos	Teacher L.	27	30%	52%	4%	15%	81%	26%
	Teacher	87	24%	31%	9%	36%	55%	15%
Other KLT materials	Teacher L.	30	40%	47%	0%	13%	87%	40%
	Teacher	85	45%	31%	5%	20%	75%	40%

Note. % Positive is the sum of % Very Helpful and % Helpful. Net Promoter is % Very Helpful minus % Not helpful at all. Teacher L. denotes "Teacher Leader."

Research Question 6: How have teacher formative assessment practices influenced student educational experience and achievement?

An evaluation of professional development in a district or school typically needs to address the multitude of mediating and moderating factors that may influence its results. Some of the mediating factors are a part of the theory that informs the design and analysis. Moderating factors however may not have been explicitly defined as elements of the intervention's logic model but may burden the evaluation by what failures were found in the execution of the intended study design and data collection plans. Specifically, we also examined how participant differences in KLT implementation may have influenced the predicted relationships and the evaluation of how implementation affected treatment comparisons. We comment on their influence on the conclusions about changes in several key relationships of the logic model[17].

---

[17] We employed a simple analysis of covariance to assess the joint influence of the implementation fidelity measures on each relationship, separately, between the mediating variables and the student outcomes.

NWEA

In our study, we considered teacher formative assessment practices as the immediate outcomes of KLT and they mediate the impact of KLT on the student's educational experiences with formative assessment use and achievement. A first step to assessing the impact of mediating factors on our outcomes was to examine how the constructs and outcomes were interrelated at baseline (before KLT) and after one year of KLT (based on KLT Cohorts 1 and 2).

Figure 18 shows both the inter-correlations among the factors posited by KLT's logic model (see Figure 1). At baseline, we found only a few of the posited relationships as expected. But, as the results in Figure 18 showed, some notable increases in correlations were suggested at the end of KLT participants' first year of implementation relative to the baseline results. For example, KLT teachers appeared to more consistently elicit evidence of student learning (T1EESL) for the purpose of adapting their instruction to meet their students' immediate learning needs (T6AISN); the correlation between the constructs increased from a baseline magnitude of 0.33 to 0.45. When a teacher more consistently sought to understand how their students learned, the students also tended to take responsibility for their own learning and to support fellow student's efforts (S1SQAL). We found that the correlation between T1EESL and S1SQAL became statistically significant (0.19) by the end of Year 1 for KLT participants in the district. More significantly, the magnitude of this relationship appeared not to be influenced by the previously observed variation in the degree to which teachers participated in required or recommended TLC activities.

We also found that, where a teacher more often identified and shared learning expectations with their students (T2ISLE), students' engagement in classroom discussions also increased (S2ENGG). This suggests that KLT had the intended impact for many KLT teachers by the end of the first intervention year. The correlation of this teacher mediator and student outcome was a strong 0.8. This relationship too remained just as strong despite the variation in aspects of KLT implementation noted earlier[18].

---
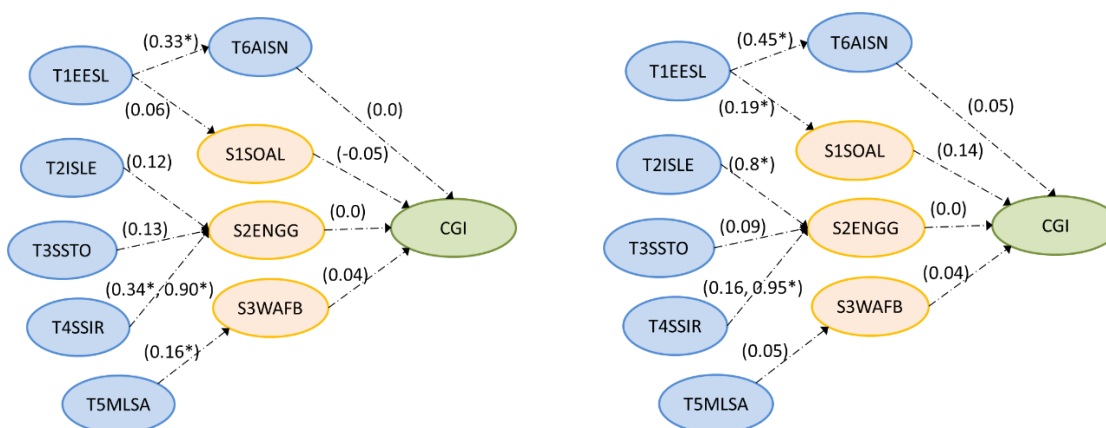
[18] An analysis of covariance is performed to detect group differences in the outcome while also controlling for differences in KLT implementation among teachers.

Figure 18. Correlations between KLT logic model factors at baseline and at the completion of one year of KLT (Please refer to Table 2 for a brief description of each construct).

Another notable finding indicated that, as viewed from the student's perspective, the more often a teacher structured opportunities to activate students as instructional resources for one another (T4SSIRb), the more students reported experiencing increased engagement in their own learning (S2ENGG). The correlation, now at 0.95, remained just about as high as it had been at baseline. One simple reason was that these two constructs shared, as suggested by theory, many of the same surveys items at its core and thus its value to the explication of the theory needs further examination. Introducing the two measures of KLT implementation, related to a teacher's participation levels in required and recommended activities, did not alter the strength of the relationship.

The only significant decrease in relationship was between how often a teacher provided feedback to move learning forward and created a structure for students to act on it (T5MLSA) and how much students reported that they welcomed and acted on feedback (S3FAWB). The correlation dropped in magnitude from 0.16, albeit small but statistically significant at baseline, to a non-statistically detectable level. This result too did not appear to be due to differences in KLT implementation.

Taken together, several of the KLT model-predicted relationships became stronger as KLT took hold in the district. We speculate that perhaps these and other remaining relationships among mediating variables and outcomes will continue to strengthen by the end of the second year of KLT. This analysis is not presently supportable however, due to design and sample size constraint.

Lastly, and turning to the more distal set of outcomes for the intervention, we found no detectable improvements in learning gains based on student mathematics and reading scores. This was not unexpected for several reasons. First, KLT is directly aimed at impacting teacher and student relationships around the values and use of formative assessments in their classrooms to facilitate success in learning. Any immediate impact on achievement and growth cannot be realistically expected. Second, Meridian is a relatively high-performing district with students averaging well above the national norms for growth, and remained so throughout the course of the two-year study. Elsewhere in our research on MAP achievement norms (NWEA, 2011), we have found that it is simply very difficult to significantly increase aggregate measures of academic performance, such as the average score for a grade-level, especially for high-performing districts.

As we discussed above, our findings and comparisons of the key constructs of KLT logic model were largely positive thus far. Several predicted relationships between teacher and student outcomes emerged and differences in implementing KLT did not appear to have influenced them. Furthermore, while we have shown that various aspects of administrative support for KLT may influence KLT implementation as indicated by a teacher's reported participation in TLC activities, the data did not indicate that administrative support affected the relationships specified in KLT's logic model.

Discussion

This study was designed to help our school district partners assess the impact of KLT. This research also offered an opportunity to address the literature by measuring the impact of formative assessment practices delivered through a professional development program like KLT. Despite research evidence which suggested that efforts to integrate the use of assessments with a curriculum might produce enhanced student (e.g., Black & Wiliam, 1998; Brookhart, 2007; Guskey, 2007; Black & Wiliam, 2007; Hattie, 2008; Shavelson et al., 2008), some had argued that the value of formative assessment practices has an acceptance level by its practitioners that may not be commensurate with the empirical evidence accumulated to date (e.g., Bennett, 2011; Briggs & Ruiz-Primo 2012; Dunn & Mulvenon, 2009; Kingston & Nash, 2011). Bennett (2011, p. 21), for example, argues "'formative assessment' is both conceptually and practically still a work-in-progress."

Our results were generally positive in that there was ample evidence to suggest that the intervention influenced classroom practices as intended. The most encouraging findings came from Meridian's changes in teacher formative assessment practices. Over time, KLT teachers reported they structure opportunities for students to take ownership of their own learning, structure opportunities to activate students as instructional resources for one another, provide feedback to move learning forward, and adapt instruction to meet students' immediate learning needs more often than control teachers and baseline measures. Both participant cohorts showed a noticeable increase in the level of accountability to their peers and to their administration to incorporate formative assessment strategies and techniques in their classroom, with Cohort 2 showing a sizable increase over baseline levels. In some instances, there were differences between participation groups. For example, Cohort 1 participants reported greater improvements in their second year on how much they identify and share learning expectations with their students which may suggest that some of KLT's intended outcomes could take longer to "kick in" or was the result of the delayed implementation in year one. Furthermore, teachers and Teacher Leaders had positive assessments about their KLT experience and reported sharing experiences with their peers in the TLCs as the most helpful aspect of the program.

Across the district, students reported that they knew what they were supposed to learn, knew how to learn, and were learning successfully throughout the duration of the study. In KLT classrooms, students reported high levels of engagement measures (e.g., involvement in classroom discussions and students welcome and act on feedback) with some positive increases over time. Furthermore, the more often a teacher identified and shared learning expectations with their students and structured opportunities to activate students as instructional resources for one another, the more students reported experiencing increased engagement in their own learning. In contrast, student achievement did not appear to be impacted by KLT. Student achievement gains (measured by the conditional growth index) were very stable over the course of the study and provide little direct evidence that KLT impacted achievement by the end of Spring 2014. There could be a number of reasons why we saw no significant changes in student achievement.

We found Meridian students perform better than national norms and this study examined aggregated student achievement and growth measures (NWEA, 2011). With a more robust sample, future studies could explore how student achievement growth may be associated with differences across schools (e.g., elementary vs. middle school) or teachers (e.g., teaching experience) or student characteristics (e.g., low vs. high performing). Alternatively, it could be too early to see measurable gains in student achievement. As many researchers postulate (e.g., Fullan, 1997), professional development programs may take time to show discernable progress in

terms of student achievement outcomes. That is, student achievement are likely lagging indicators of positive changes in the classroom. We hope to continue our partnership with Meridian and track student growth in subsequent years to explore this possibility.

As we have indicated earlier, one question that interests our partners is of course how KLT works. The answer to this question must begin with a clearly specified logic model that has guided its evaluation. From the logic model, we are able to develop several survey instruments to collect the data we needed to test the relationships predicted by the model. The rudimentary comparisons of baseline results with that from KLT participants at the end of their first year provided evidence that formative assessment practices in KLT classrooms have either emerged or improved together as predicted, all despite some difference in implementation fidelity.

Lastly, we found that although administrative support for KLT had tended to increase a teacher's reported participation in TLC activities, the results did not indicate that administrative support or implementation fidelity served to modify our conclusions about key relationships in KLT's logic model measured at Meridian.

Challenges for Research in Professional Development

Data collected for this study focused primarily on teachers' formative assessment practices, students' educational experience, and student achievement measures. Although teacher and student surveys yielded a considerable amount of data, our evolving study design and the shifting study population posed several limitations. We summarize the major considerations below.

After defining the initial study population, several factors impacted the resulting study participation subgroups. For instance, the district made the decision to expand KLT in additional schools (primarily as part of Title I improvement plans) which depleted our control school sample and changed the composition of our participant group in potentially meaningful ways. While this decision in and of itself was a positive signal that the district saw justification to expand KLT's implementation, it did result in weakened data and significantly limited our scope for inferences and the statistical power of conventional analyses.

Given the data realized, we opted to employ a series of comparisons bearing in mind two confounding factors. We needed to deal with possible compositional effects due to known changes in (cross-sectional) teacher groups over time – a complex mix of attrition and selection (including contamination). For example, for any single outcome, confidence in an inference increased when major subgroups showed corroborative patterns. In the case of S2ENGG, for example, we inferred the following: In the base year (2012), all three participation subgroups do not differ in the levels of student-reported engagement but, for those students in KLT classrooms, student-reported engagement levels increased by about a half of a standard deviations above base levels. Generally, we discounted estimates with large standard errors. We applied $1.4 \times SE$ to each side of a plotted mean estimate as a visual comparison interval, in which overlapping error bars would signal that means are statistically indistinguishable. Furthermore, this positive result did not appear to be due to differences in the number of data elements within participation patterns (we "discount" the contribution of information with large standard errors to an aggregate, see Figure 4). This statistical rationale guided the use of the available data and our interpretation of the results. We also did not address the presence of bias due to selection and merely took the standard errors of cross-sectional mean estimates into account in our within and between participation subgroup comparisons. Potentially, more complex cutting-edge statistical remedies appear to be available in the work of Hong (2010).

However, it should be noted that such sparse data, with strong attrition as the study progressed, would severely limit the potential of the new methodologies such as Hong (2010).

In terms of a more comprehensive framework for studying the impact of intervention in schools, prominent urban education scholar and Carnegie Foundation for the Advancement of Teaching President Anthony Bryk and his colleagues recently argued that educational research has much in common with the public health model. He suggested that randomized control trials are often unrealistic and we need an accumulating evidence strategy. This naturalist approach, under the banner of "improvement research", is anchored in theory about advancing improvement reliably on a large scale. To accomplish this mission, researchers must "assess (even crudely) the value added to learning associated with individual classrooms and schools and investigating what might be driving observed variability in these effects" (Bryk, 2010, p. 22). Future research on KLT, or other interventions, can certainly draw upon this framing of improvement research with benefit.

The limitations above notwithstanding, we hope the approach we developed helps Meridian, and other partner districts, assess the value of the investment in KLT, as well as similar professional development opportunities. Furthermore, the findings from this study will inform recommendations to improve the KLT program and its delivery, and extend our understanding of formative assessment use in classrooms and under what conditions they maximize chances of improved student engagement and learning.

References

Babbie, E. (1990). *Survey research methods*, *second edition*. Belmont, CA: Wadsworth Publishing.

Ball, D. L. (1996). Teacher learning and the mathematics reforms: What we think we know and what we need to learn. *Phi Delta Kappan*, *77*(7), 500-8.

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18,* 5 – 25.

Bergan, J. R., Sladeczek, I. E., Schwarz, R. D., & Smith, A. N. (1991). Effects of a measurement and planning system on kindergartners' cognitive development and educational programming. *American Educational Research Journal*, *28*(3), 683-714.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.

Black, P., & Wiliam, D. (2003). In praise of educational research: Formative assessment. *British Educational Research Journal,* 29, 623– 637.

Black, P., & Wiliam, D. (2007). *A theoretical foundation for formative assessment*. In J. H. McMillan (Ed.), Formative classroom assessment: Theory into practice. (pp. 29–42). New York: Teachers College Press.

Black, P., & Wiliam, D. (2009). Developing a theory of formative assessment. *Educational Assessment, Evaluation and Accountability,* 21, 5–31.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Kappan Classic*, 92(1), 81-90.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Brace, I. (2004). Questionnaire design: How to plan, structure and write survey material for effective market research. London: *Market Research in Practice Series.*

Briggs, D.C. & Ruiz-Primo, M.A. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice,* 31(4), 13-17.

Brookhart, S. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 43–62). New York: Teachers College Press.

Brookhart, S., Moss, C., & Long, B. (2008). Formative assessment that empowers. *Educational Leadership*, 66, 52–57.

Bryk, A. (2010). Systematic Naturalistic Inquiry: Toward a Science of Performance Improvement. *Society for Research on Education Effectiveness*. Washington D.C. March, 2010.

Butler, D. L., Lauscher, H. N., Jarvis-Selinger, S., & Beckingham, B. (2004). Collaboration and self-regulation in teachers' professional development. *Teaching and teacher education*, *20*(5), 435-455.

Cauley, K. M. & McMillan, J.H. (2010). Formative assessment techniques to support student motivation and achivment. *Clearning House: A Jounral of Educational Strategies*, 83(1), 1-6.

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *83*(1), 1-6.

Cech, S. (2007). Test industry split over 'formative' assessment. *Edweek,* 28, no. 4: 1, 15.

Century, J., Rudnick, M. & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, 31(2), 199-218.

Child, D. (1990). *The essentials of factor analysis, second edition*. London: Cassel Educational Limited.

Cochran-Smith, M., & Lytle, S. L. (1999). Relationships of knowledge and practice: Teacher learning in communities. *Review of research in education*, 249-305.

Cordray, D. S., & Jacobs, N. (2005, June). *Treatment fidelity and core components in school-based ATOD prevention programs.* Paper presented at the Annual Meeting of the Society for Prevention Research, Washington, DC.

Craighead, W.E. & Nemeroff, C. B. (2004). *The Concise Corsini Encyclopedia of Psychology and Behavioral Science*. Hoboken, NJ: John Wiley & Sons Inc.

Darling-Hammond, L. (1996). The Quiet Revolution: Rethinking Teacher Development. *Educational leadership*, *53*(6), 4-10.

Dunn, K. & Mulvenon, S. (2009). A Critical Review of Research on Formative Assessment: The Limited Scientific Evidence of the Impact of Formative Assessment in Education. *Practical Assessment, Research, & Evaluation,* 14(7), 1-11.

Educational Testing Service. (2009). Research rationale for the Keeping Learning on Track program. Princeton: Author.

Educational Testing Service. (2010). *About the KLT program.* Princeton, NJ: Author.

Fernandes, M., & Fontana, D. (1996). Changes in the control beliefs in Portuguese primary school pupils as a consequence of employment of self-assessment strategies. *British Journal of Educational Psychology,* 66, 301-313.

Finn, J. D., & Rock, D. A., (1997). Academic success among students at risk for school failure. *The Elementary School Journal,* 95(5), 421-434.

Fuch, L. S., & Fuch, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53,* 199-208.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal,* 28(3), 617-641.

Fullan, M. (1983). Evaluating program implementation: What can be learned from follow through. *Curriculum Inquiry*, 13(2), 215-227.

Fullan, M. (1997). Emotion and hope: Constructive concepts for complex times. In A. Hargreaves (Ed*.), Rethinking educational change with heart and mind*. Alexandria, VA: Association for Supervision and Curriculum Development.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American educational research journal*, *38*(4), 915-945.

Goldstein , H., & Healy , M. J. R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society*, Series A, 158, 175-177.

Grunwald, G.A. (2012). *For every child, multiple measures*. NWEA.

Guskey, T. R. (2007). Formative classroom assessment and Benjamin S. Bloom: Theory, research, and practice. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 63–78). New York: Teachers College Press.

Hall, G. E., & Loucks, S. F. (1978). *Innovation Configurations: Analyzing the Adaptations of Innovations*. National Institute of Education, Washington D.C.

Hattie, J. A. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement.* New York, NY: Routledge.

Hong, G. (2010). Marginal mean weighting through stratification: Adjustment for selection bias in multilevel data. *Journal of Educational and Behavioral Statistics*, 35, 499-531.

Hulleman, C., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Educational Effectiveness, 2*(1), 88–110.

Iarossi, G. (2006). *The power of survey design: A user's guide for managing surveys, interpreting results, and influencing respondents.* Washington, D.C.: The World Bank.

Jones, E. D., & Krouse, J. P. (1988). The effectiveness of data-based instruction by student teachers in classrooms for pupils with mild learning handicaps. *Teacher Education and Special Education, 1*(1), 9-19.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL7, a Guide to the Program and Applications.* SPSS Publications.

Kingston, N. & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28-37.

Kingston, N. & Nash, B. (2012). How many formative assessment angels can dance on the head of a meta-analytic pin: .2. *Educational Measurement: Issues and Practice, 31*(4), 18-19.

Loucks-Horsley, S., Stiles, K. E., Mundry, M. S. E., Love, N. B., & Hewson, P. W. (2009). *Designing professional development for teachers of science and mathematics*. Corwin Press.

Martinez, J. G. R., & Martinez, N. C. (1992). Re-examining repeated testing and teacher effects in a remedial mathematics course. *British Journal of Educational Psychology*, 62, 356-363.

McCoach, D. B., Gable, R. K., & Madura, J. (2013). *Instrument development in the affective domain*. Springer-Verlag New York Incorporated.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199–218.

Nolen, S.B. (2011). The role of educational systems in the link between formative assessment and motivation. *Theory Into Practice, 50*, 319-326.

Northwest Evaluation Association. (2011). *RIT scale norms: For use with Measures of Academic Progress (MAP®) and MAP® for Primary Grades.* Portland, OR: Author.

Peterson, P. L., Carpenter, T., & Fennema, E. (1989). Teachers' knowledge of students' knowledge in mathematics problem solving: Correlational and case analysis. *Journal of Educational Psychology, 81*(4), 558-569.

Popham, W. J. (2008). *Transformative assessment.* Alexandria, VA: ASCD.

Raudenbush, S.W. & Bryk, S.B. (2001). *Hierarchical linear models: Applications and data analysis methods (advanced quantitative techniques in the social sciences).* Thousand Oaks, CA: Sage Publications, Inc.

Reise. S. P., Ventura, J., Nuechterlein, K. H., & Kim, K. H. (2005). An illustration of multilevel factor analysis. *Journal of Personality Assessment*, 84, 126—36.

Ruston, A. (2005). Formative assessment: a key to deep learning? *Medical Teacher, 27*(6), 509-513.

Saettler, P. (1990). *The Evolution of American Educational Technology.* Englewood, Colorado: Libraries Unlimited, Inc.

Saunders, W. M., Goldenberg, C. N., & Gallimore, R. (2009). Increasing achievement by focusing grade-level teams on improving classroom learning: A prospective, quasi-experimental study of Title I schools. *American Educational Research Journal*, *46*(4), 1006-1033.

Saxe, G., Gearhart, M., & Nasir, N. S. (2001). Enhancing students' understanding of mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education*, 4, 55–79.

Scriven, M. (1967). *The methodology of evaluation.* In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), Perspectives of curriculum evaluation, 39-83. Chicago, IL: Rand McNally.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M. K., & Yin, Y. (2008). On the Impact of Curriculum-Embedded Formative Assessment on Learning: A Collaboration between Curriculum and Assessment Developers, *Applied Measurement in Education*, 21, 295-314

Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. *Journal of Educational Psychology,* 82(1), 22-32.

Sparks, D., & Loucks-Horsley, S. (1989). Five models of staff development for teachers. *Journal of staff development*, *10*(4), 40-57.

Stiggins, R. J. (2002). Assessment Crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, 83(10), 758-765.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of educational change*, *7*(4), 221-258.

Supovitz, J. A., & Christman, J. B. (2003). Developing Communities of Instructional Practice: Lessons from Cincinnati and Philadelphia. CPRE Policy Briefs. RB-39. *Consortium for Policy Research in Education*.

Thum, Y. M. (2014, in press). The Effective Use of Student and School Descriptive Indicators of Learning Progress: From the Conditional Growth Index to the Learning Productivity Measurement System. In R. Lissitz (Ed.) *Value Added Modeling and Growth Modeling with Particular Application to Teacher and School Effectiveness* (pages). Maple Grove, MN: JAM Press.

Thum, Y. M., & Bhattacharya, S. K. (2001). Detecting a change in school performance: A Bayesian analysis for a multilevel join point problem. *Journal of Educational and Behavioral Statistics*, 26, 443-468.

Thum, Y. M., & Xiang, Y. (2010). *Representing the Learning Productivity of Students at a School with an Effect-Size Indicator in a Pay-for-Performance Application.* Unpublished paper. Portland, OR: Author.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and teacher education*, *24*(1), 80-91.

Wang, M.C., Nojan, M., Strom, C.D. & Walberg, H.J. (1984). The utility of degree of implementation measures in program implementation and evaluation research. *Curriculum Inquiry*, 14(3), 249-286.

Weiss, I. R., Montgomery, D. L., Ridgway, C. J., & Bond, S. L. (1998). *Local systemic change through teacher enhancement: Year three cross-site report.* Horizon Research Inc., Chapel Hill, NC.

Whiting, B., Van Burgh, J. W., & Render, G F. (1995). *Mastery learning in the classroom.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Wiliam. D. (2012, July). *Why formative assessment needs to be a priority for every school*, Workshop presentation, Boise, ID.

Appendix A: Tables and Figures

Table A1
Percentage of Respondents to Teacher Survey Items.

1. When I hear "Formative Assessment" I think it includes

Item ID	N	Yes	No		More* Positive	Item
		1	0			
EC01a	315	56.51	43.49		n.s.	Gathering minute by minute, day by day evidence of learning
EC01b	315	42.54	57.46		n.s.	Giving the same tests or quizzes to students across different classrooms
EC01c	315	62.86	37.14		n.s.	Students reviewing their own work against rubrics or other standards of performance
EC01d	315	79.05	20.95		n.s.	Teachers providing students with feedback about their work
EC01e	315	54.92	45.08		n.s.	Peers providing students with feedback about their work

2. Please indicate how strongly you agree or disagree with each statement below

Item ID	N	Strongly Disagree		Agree		Strongly Agree	More Positive	Item
		1	2	3	4	5		
EC02a	315	10.79	5.08	6.98	12.7	64.44	n.s.	It is important for students to have a strong ownership and understanding of where they are now, where they need to go, and how to get there.
EC02b	315	7.62	10.79	17.78	22.22	41.59	n.s.	Formative assessment is valuable for determining whether students have a deep understanding of content.
EC02c	315	11.43	4.44	11.75	15.87	56.51	n.s.	Formative assessment is valuable for improving teaching.
EC02d	314	10.83	5.1	9.87	17.52	56.69	n.s.	Formative assessment is valuable for improving learning.

NWEA.

4. Please tell us about your current instructional practices

| Item ID | N | Never | Occasionally | | | Always | More Positive | Item |
		1	2	3	4	5		
EC04a	296	0	0.68	17.91	59.8	21.62	n.s.	I identify and share the intended learning and success criteria with students.
EC04b	296	0	0	3.04	58.78	38.18	CTL	I provide instruction that is closely in tune with the learning needs of my students.
EC04c	295	0	1.69	21.69	56.95	19.66	n.s.	I have students articulate their thinking exposing their misconceptions.
EC04d	296	0	2.03	14.86	60.14	22.97	n.s.	I provide specific feedback to students that tell them what must be done to improve.
EC04e	295	0	0.68	13.56	54.24	31.53	n.s.	I involve all students in classroom discussions, thinking about the questions, understanding and acting on the evidence of learning.
EC04f	295	0.68	5.76	47.12	39.32	7.12	n.s.	I collect responses from all students simultaneously.

5. Indicate how frequently you do the following practices

| Item ID | N | Never | Occasionally | | | Always | More Positive | Item |
		1	2	3	4	5		
EC05a	294	0	9.18	40.82	46.6	3.4	n.s.	I provide the structures for students to reflect on and formatively use the feedback to improve their understanding.
EC05b	294	0	8.45	32.09	54.05	5.41	n.s.	I provide students the structured opportunity for them to act as instructional resources for each other.

6. I often intentionally design opportunities for students to

| Item ID | N | Never | Occasionally | | | Always | More Positive | Item |
		1	2	3	4	5		
EC06a	295	1.36	15.25	37.63	39.66	6.1	n.s.	Reflect on whether they have learned something well enough.
EC06b	294	0.34	20.07	37.41	36.73	5.44	n.s.	Reflect on how they learned something.
EC06c	294	3.06	21.77	49.32	24.49	1.36	n.s.	Choose how they want to learn something.

48

NWEA.

7. Please indicate how strongly you agree or disagree with each statement below

Item ID	N	Strongly Disagree 1	2	Agree 3	4	Strongly Agree 5	More Positive	Item
EC07a	296	1.69	18.24	44.93	25	10.14	CTL	I understand a significant number of formative assessment strategies and techniques I can use in the classroom.
EC07b	296	1.69	17.91	37.84	26.35	16.22	CTL	I am involved in a rigorous process for planning changes to my current teaching practice.
EC07c	293	9.56	34.81	32.76	17.75	5.12	n.s.	I receive ongoing support to improve my use of formative assessment strategies and techniques.
EC07d	296	6.76	22.64	37.5	22.64	10.47	n.s.	I am accountable to my peers to incorporate formative assessment strategies and techniques into my classroom.
EC07e	294	2.72	15.99	40.82	26.19	14.29	n.s.	I am accountable to my administrators to incorporate formative assessment strategies and techniques into my classroom.

9. How important is it for student to have opportunities to

Item ID	N	Not at All 1	2	Moderately 3	4	Very 5	More Positive	Item
EC09a	156	0	0.64	10.9	41.03	47.44	n.s.	Reflect on whether they have learned something well enough.
EC09b	156	0	1.28	14.74	44.87	39.1	n.s.	Reflect on how they learned something.
EC09c	156	0	3.85	27.56	47.44	21.15	n.s.	Choose how they want to learn something.

10. Please indicate how strongly you agree or disagree with each statement below

Item ID	N	Strongly Disagree 1	2	Agree 3	4	Strongly Agree 5	More Positive	Item
EC010a	314	1.59	11.78	37.9	30.57	18.15	n.s.	Formative assessment is a high priority for my district.
EC010b	314	1.59	7.96	35.67	31.21	23.57	n.s.	Formative assessment is a high priority for my school.
EC010c	313	0.32	4.15	24.6	37.38	33.55	n.s.	Formative assessment is a high priority for me.

NWEA.

Either KLT® teachers, or Control (CTL) teachers, or neither (n.s.), show more positive response overall on items in the teacher surveys, based on a chi-square tests for differences in frequency distributions of teacher responses and at the nominal $\alpha=0.05$ level. "n.s." denotes "not statistically significant."

Table A2
Summary of Factor Analyses Results.

	S1SOAL	S2ENGG	S3WAFB	T0VAFA	T1EESL	T2ISLE	T3SSTO	T4SSIR*	T5MLSA	T6AISN	T7PDVA
n observations	254	254	254	442	NA	230	378	163	378	377	378
n Factors	1	1	1	1	NA	1	1	2	1	1	1
Cronbach's α	0.746	0.638	0.834	0.958	NA	0.482	0.643	NA	0.673	0.680	0.818
Prop. of Variance	0.664	0.580	0.464	0.914	NA	0.586	0.570	0.384	0.584	0.475	0.581
					Descriptive Statistics						
Factor 1: *n* of Items	3	3	8	3	NA	3	4	2	4	5	5
Factor 1: Mean	0.065	0.060	0.065	0.000	NA	0.000	0.000	0.125	0.000	0.000	0.000
Factor 1: *SD*	0.257	.0353	0.038	1.000	NA	1.000	1.000	0.752	1.000	1.000	1.000
Factor 2: *n* of Items								4			
Factor 2: *M*								0.039			
Factor 2: *SD*								0.288			
Factor Correlation								0.234			

Table A3

Correlations Between Selected Survey Items and Measures of KLT Logic Model Constructs.

	y1_CGI	y2_CGI	S1SOAL	S2ENGG	S3FAWB	T0VAFA	ec04f	SEC02d*	T2ISLE	T3SSTO	T4SSIRa	T4SSIRb	T5MLSA	T6AISN
y1_CGI	1 574													
y2_CGI	0.364 <.0001 486	1 613												
S1SOAL	-0.049 0.571 139	0.015 0.859 152	1 197											
S2ENGG	0.004 0.963 139	0.032 0.697 152	0.622 <.0001 197	1 197										
S3FAWB	0.038 0.660 139	0.071 0.386 152	0.689 <.0001 197	0.922 <.0001 197	1 197									
T0VAFA	-0.058 0.430 190	0.103 0.133 212	0.149 0.076 144	0.233 0.005 144	0.250 0.003 144	1 311								
ec04f	-0.061 0.403 191	0.085 0.216 215	0.057 0.497 144	0.028 0.744 144	0.038 0.655 144	0.127 0.030 291	1 295							
SEC02d*	0.046 0.594 139	0.084 0.301 152	0.544 <.0001 197	0.804 <.0001 197	0.751 <.0001 197	0.268 0.001 144	0.078 0.351 144	1 197						
T2ISLE	-0.016 0.854 134	-0.057 0.490 147	0.111 0.135 183	0.117 0.115 183	0.140 0.059 183	0.097 0.271 131	-0.039 0.654 131	0.049 0.514 183	1 183					
T3SSTO	0.007 0.923 186	-0.030 0.666 208	0.126 0.135 142	0.133 0.115 142	0.167 0.047 142	0.043 0.470 287	0.181 0.002 287	0.110 0.193 142	0.036 0.684 129	1 288				
T4SSIRa	-0.071 0.477 102	0.102 0.293 108	0.164 0.059 134	0.335 <.0001 134	0.329 0.000 134	0.206 0.017 134	0.227 0.008 134	0.265 0.002 134	-0.067 0.449 130	0.596 <.0001 132	1 134			
T4SSIRb	-0.015 0.883 102	0.071 0.468 108	0.570 <.0001 134	0.896 <.0001 134	0.917 <.0001 134	0.255 0.003 134	0.029 0.739 134	0.621 <.0001 134	0.096 0.279 130	0.239 0.006 132	0.499 <.0001 134	1 134		
T5MLSA	0.006 0.933 186	-0.034 0.629 208	0.132 0.117 142	0.114 0.177 142	0.162 0.054 142	-0.006 0.924 287	0.226 0.000 288	0.092 0.278 142	0.104 0.240 129	0.866 <.0001 286	0.573 <.0001 132	0.239 0.006 132	1 288	
T6AISN	-0.003 0.964 186	-0.014 0.842 208	0.132 0.118 142	0.132 0.117 142	0.168 0.046 142	0.065 0.273 286	0.337 <.0001 287	0.119 0.158 142	0.028 0.753 129	0.987 <.0001 287	0.612 <.0001 132	0.234 0.007 132	0.865 <.0001 286	1 287
T7PDVA	-0.090 0.222 186	-0.016 0.823 207	0.019 0.818 143	-0.017 0.843 143	-0.023 0.782 143	0.002 0.970 288	0.120 0.042 287	0.007 0.932 143	0.081 0.360 130	0.362 <.0001 284	0.287 0.001 133	0.061 0.483 133	0.403 <.0001 283	0.364 <.0001 283

Note. * Classroom mean of student survey items.

NWEA.

Appendix B: Teacher, Teacher Leader, and Administrator Survey

Spring 2014 Keeping Learning on Track Participant Survey

Confidentiality Statement:

Dear Educator,

Thank you for taking the time to complete this survey. Your responses will help Northwest Evaluation Association (NWEA) understand the conditions under which various approaches to professional development in formative assessment strategies and techniques improve student learning. NWEA has strict guidelines regarding the confidentiality of data, including:

• No responses collected via this survey will be reported so that they can be attributed to any one person.
• Your district personnel will not have access to individual responses.
• An individual's survey responses are combined with those of many others and used in NWEA's statistical analysis supporting this research.

This 15-20 minute voluntary survey does require you to record personal information, like your name, so that it is possible for NWEA to combine response data with subsequent surveys. A summary report will be made available to the interested public at http://www.nwea.org.

If you have any questions about NWEA confidentiality and procedures, please contact: Dr. Yeow Meng Thum, Principal Investigator (503) 624-1951.

Thank you very much.
NWEA Research Team

Spring 2014 Keeping Learning on Track Participant Survey

Formative Assessment

Please tell us about your knowledge of formative assessment.

1. Please indicate how long you have been a participant in KLT.

○ One year

○ Two years

Other (please specify)

[   ]

2. When I hear "Formative Assessment" I think it includes: (select all that apply)

☐ Gathering minute by minute, day by day evidence of learning

☐ Giving the same tests or quizzes to students across different classrooms

☐ Students reviewing their own work against rubrics or other standards of performance

☐ Teachers providing students with feedback about their work

☐ Peers providing students with feedback about their work

☐ Other (please specify)

[   ]

As you answer the questions below, please evaluate your responses based on KLT's definition of formative assessment: Students and teachers continuously using evidence of learning to adapt what happens in the classroom.

3. Please indicate how strongly you agree or disagree with each statement below.

	Strongly Disagree		Agree		Strongly Agree
It is important for students to have a strong ownership and understanding of where they are now, where they need to go, and how to get there.	○	○	○	○	○
Formative assessment is valuable for determining whether students have a deep understanding of content.	○	○	○	○	○
Formative assessment is valuable for improving teaching.	○	○	○	○	○
Formative assessment is valuable for improving learning.	○	○	○	○	○

NWEA

Spring 2014 Keeping Learning on Track Participant Survey

Position/Title

4. Please select which position/title best describes you.

◯ Administrator

◯ Teacher Leader

◯ Teacher

◯ Other (please specify)

[   ]

NWEA.

Spring 2014 Keeping Learning on Track Participant Survey

TLC Leader Current Instructional Practices

Please tell us about your current instructional practices.

5. Indicate how frequently you do the following practices.

	Never	Seldom	Occasionally	Frequently	Always
I identify and share the intended learning and success criteria with students.	○	○	○	○	○
I provide instruction that is closely in tune with the learning needs of my students.	○	○	○	○	○
I have students articulate their thinking exposing their misconceptions.	○	○	○	○	○
I provide specific feedback to students that tell them what must be done to improve.	○	○	○	○	○
I involve all students in classroom discussions, thinking about the questions, understanding and acting on the evidence of learning.	○	○	○	○	○
I collect responses from all students simultaneously.	○	○	○	○	○

6. Indicate how frequently you do the following practices.

	Never	Seldom	Occasionally	Frequently	Always
I provide the structures for students to reflect on and formatively use feedback to improve their understanding.	○	○	○	○	○
I provide students the structured opportunity for them to act as instructional resources for each other.	○	○	○	○	○

NWEA.

7. I often intentionally design opportunities for students to:

	Never	Seldom	Occasionally	Frequently	Always
Reflect on whether they have learned something well enough.	◯	◯	◯	◯	◯
Reflect on how they learned something.	◯	◯	◯	◯	◯
Choose how they want to learn something.	◯	◯	◯	◯	◯

57

NWEA

TLC Leader Professional Development

8. Please indicate how strongly you agree or disagree with each statement below.

	Strongly Disagree		Agree		Strongly Agree
I understand a significant number of formative assessment strategies and techniques I can use in the classroom.	○	○	○	○	○
I am involved in a rigorous process for planning changes to my current teaching practice.	○	○	○	○	○
I receive ongoing support to improve my use of formative assessment strategies and techniques.	○	○	○	○	○
I am accountable to my peers to incorporate formative assessment strategies and techniques into my classroom.	○	○	○	○	○
I am accountable to my administrators to incorporate formative assessment strategies and techniques into my classroom.	○	○	○	○	○

NWEA.

TLC Leader Organizational Priorities

9. Please indicate how strongly you agree or disagree with each statement below.

	Strongly Disagree		Agree		Strongly Agree
Formative assessment is a high priority for my district.	◯	◯	◯	◯	◯
Formative assessment is a high priority for my school.	◯	◯	◯	◯	◯
Formative assessment is a high priority for me.	◯	◯	◯	◯	◯

NWEA.

TLC Leader KLT Implementation

Please tell us about the implementation of KLT.

10. My Teacher Learning Community (TLC) completed _____ meetings this year.

[   ]

11. Our typical meeting lasts_____ minutes.

[   ]

12. My TLC completed _____ modules this academic year.

[   ]

13. Indicate how frequently you, your TLC, or your administration does the following:

	Never	Seldom	Occasionally	Frequently	Always
I attend my KLT TLC meetings.	○	○	○	○	○
I develop a Personal Action Plan at each TLC meeting.	○	○	○	○	○
I implement my Action Plans.	○	○	○	○	○
My TLC follows the module sequence.	○	○	○	○	○
My school administration provides me with the necessary support for my TLC (e.g., dates, times, materials).	○	○	○	○	○

Spring 2014 Keeping Learning on Track Participant Survey

14. Indicate approximately how frequently you or your administration does the following:

	Never	Once a Year	Once a Quarter	Once a Month	Once a Week
I share lesson plans.	○	○	○	○	○
I ask for and offer feedback to support others in embedding formative assessment in their teaching.	○	○	○	○	○
I watch a peer implement formative assessment techniques.	○	○	○	○	○
Peer(s) watch me implement formative assessment techniques.	○	○	○	○	○
My school administration checks in with me regarding where my TLC is in learning (what module) and what support I might need.	○	○	○	○	○

15. In your opinion, please indicate how strongly you agree or disagree with the statement below.

	Strongly Disagree		Agree		Strongly Agree
My school administration is familiar with the KLT materials and able to recommend supports to teachers and TLC Teacher Leaders.	○	○	○	○	○

NWEA

Spring 2014 Keeping Learning on Track Participant Survey

Teacher Leader KLT Evaluation

16. Please tell us about your KLT experience by indicating how helpful each aspect of KLT was to you:

	Not helpful at all		Helpful		Very helpful	N/A
Foundations Day	○	○	○	○	○	○
Monthly KLT TLC meetings	○	○	○	○	○	○
Developing Personal Action Plans	○	○	○	○	○	○
Sharing experiences with peers in TLCs	○	○	○	○	○	○
KLT modules used in TLCs	○	○	○	○	○	○
KLT Videos	○	○	○	○	○	○
Other KLT materials	○	○	○	○	○	○

NWEA

Spring 2014 Keeping Learning on Track Participant Survey

Teacher Current Instructional Practices

Please tell us about your current instructional practices.

17. Indicate how frequently you do the following practices.

	Never	Seldom	Occasionally	Frequently	Always
I identify and share the intended learning and success criteria with students.	○	○	○	○	○
I provide instruction that is closely in tune with the learning needs of my students.	○	○	○	○	○
I have students articulate their thinking exposing their misconceptions.	○	○	○	○	○
I provide specific feedback to students that tell them what must be done to improve.	○	○	○	○	○
I involve all students in classroom discussions, thinking about the questions, understanding and acting on the evidence of learning.	○	○	○	○	○
I collect responses from all students simultaneously.	○	○	○	○	○

18. Indicate how frequently you do the following practices.

	Never	Seldom	Occasionally	Frequently	Always
I provide the structures for students to reflect on and formatively use feedback to improve their understanding.	○	○	○	○	○
I provide students the structured opportunity for them to act as instructional resources for each other.	○	○	○	○	○

NWEA.

19. I often intentionally design opportunities for students to:

	Never	Seldom	Occasionally	Frequently	Always
Reflect on whether they have learned something well enough.	○	○	○	○	○
Reflect on how they learned something.	○	○	○	○	○
Choose how they want to learn something.	○	○	○	○	○

NWEA

Spring 2014 Keeping Learning on Track Participant Survey

Teacher Professional Development

20. Please indicate how strongly you agree or disagree with each statement below.

	Strongly Disagree		Agree		Strongly Agree
I understand a significant number of formative assessment strategies and techniques I can use in the classroom.	○	○	○	○	○
I am involved in a rigorous process for planning changes to my current teaching practice.	○	○	○	○	○
I receive ongoing support to improve my use of formative assessment strategies and techniques.	○	○	○	○	○
I am accountable to my peers to incorporate formative assessment strategies and techniques into my classroom.	○	○	○	○	○
I am accountable to my administrators to incorporate formative assessment strategies and techniques into my classroom.	○	○	○	○	○

Spring 2014 Keeping Learning on Track Participant Survey

Teacher Organizational Priorities

21. Please indicate how strongly you agree or disagree with each statement below.

	Strongly Disagree		Agree		Strongly Agree
Formative assessment is a high priority for my district.	◯	◯	◯	◯	◯
Formative assessment is a high priority for my school.	◯	◯	◯	◯	◯
Formative assessment is a high priority for me.	◯	◯	◯	◯	◯

NWEA.

Teacher KLT Implementation

Please tell us about the implementation of KLT.

22. My Teacher Learning Community (TLC) completed _____ meetings this year.

[   ]

23. Our typical meeting lasts_____ minutes.

[   ]

24. My TLC completed _____ modules this academic year.

[   ]

25. Indicate how frequently you, your TLC, or your administration does the following:

	Never	Seldom	Occasionally	Frequently	Always
I attend my KLT TLC meetings.	○	○	○	○	○
I develop a Personal Action Plan at each TLC meeting.	○	○	○	○	○
I implement my Action Plans.	○	○	○	○	○
I co-lead TLC meetings.	○	○	○	○	○
My school administration provides me with the necessary support for my TLC (e.g., dates, times, materials).	○	○	○	○	○

26. Indicate approximately how frequently you or your administration does the following:

	Never	Once a Year	Once a Quarter	Once a Month	Once a Week
I share lesson plans.	○	○	○	○	○
I ask for and offer feedback to support others in embedding formative assessment in their teaching.	○	○	○	○	○
I watch a peer implement formative assessment techniques.	○	○	○	○	○
Peer(s) watch me implement formative assessment techniques.	○	○	○	○	○
My school administration checks in with me regarding where my TLC is in learning (what module) and what support I might need.	○	○	○	○	○

NWEA

27. In your opinion, please indicate how strongly you agree or disagree with the statement below.

	Strongly Disagree		Agree		Strongly Agree
My school administration is familiar with the KLT materials and able to recommend supports to teachers and TLC Teacher Leaders.	◯	◯	◯	◯	◯

NWEA

Teacher KLT Evaluation

28. Please tell us about your KLT experience by indicating how helpful each aspect of KLT was to you:

	Not helpful at all		Helpful		Very helpful	N/A
Foundations Day	○	○	○	○	○	○
Monthly KLT TLC meetings	○	○	○	○	○	○
Developing Personal Action Plans	○	○	○	○	○	○
Sharing experiences with peers in TLCs	○	○	○	○	○	○
KLT modules used in TLCs	○	○	○	○	○	○
KLT Videos	○	○	○	○	○	○
Other KLT materials	○	○	○	○	○	○

NWEA

Spring 2014 Keeping Learning on Track Participant Survey

TLC Leader/Teacher Demographic Background

Please tell us a little more about yourself.

29. Select which category best describes you. What subjects do you currently teach? (select all that apply)

☐ General elementary

☐ English/Language Arts

☐ Mathematics

☐ Social Studies/Social Science

☐ History

☐ Science

☐ Foreign Languages

☐ Physical Education/Health

☐ English as a Second Language

☐ Special Education

☐ Gifted

☐ Art/Music

☐ Drama/Theater

☐ Business/Marketing

☐ Vocational/Technical

Other (please specify)

[   ]

NWEA.

Administrator/Other Current Instructional Practices

In your opinion, please indicate the importance of each instructional practice.

30. How important is it for teachers to:

	Not at All Important	Of Little Importance	Moderately Important	Important	Very Important
Identify and share the intended learning and success criteria with students.	◯	◯	◯	◯	◯
Provide instruction that is closely in tune with the learning needs of students.	◯	◯	◯	◯	◯
Have students articulate their thinking exposing their misconceptions.	◯	◯	◯	◯	◯
Provide specific feedback to students that tell them what must be done to improve.	◯	◯	◯	◯	◯
Involve all students in classroom discussions, thinking about the questions, understanding and acting on the evidence of learning.	◯	◯	◯	◯	◯
Collect responses from all students simultaneously.	◯	◯	◯	◯	◯
Provide the structures for students to reflect on and formatively use feedback to improve their understanding.	◯	◯	◯	◯	◯
Provide students the structured opportunity for them to act as instructional resources for each other.	◯	◯	◯	◯	◯

31. How important is it for student to have opportunities to:

	Not at All Important	Of Little Importance	Moderately Important	Important	Very Important
Reflect on whether they have learned something well enough.	◯	◯	◯	◯	◯
Reflect on how they learned something.	◯	◯	◯	◯	◯
Choose how they want to learn something.	◯	◯	◯	◯	◯

NWEA

Administrator/Other Organizational Priorities

32. Please indicate how strongly you agree or disagree with each statement below.

	Strongly Disagree		Agree		Strongly Agree
Formative assessment is a high priority for my district.	○	○	○	○	○
Formative assessment is a high priority for my school.	○	○	○	○	○
Formative assessment is a high priority for me.	○	○	○	○	○

NWEA.

Administrator/Other KLT Implementation

33. Concerning the implementation of KLT, please indicate how often you provide KLT Teacher Leaders with the necessary support for their Teacher Learning Communities (TLCs). e.g., dates, times, materials.

Never	Seldom	Occasionally	Frequently	Always
○	○	○	○	○

34. Indicate approximately how frequently you do the following:

	Never	Once a Year	Once a Quarter	Once a Month	Once a Week
I check in with TLC Teacher Leaders regarding where they are in their learning (what module) and what support they might need.	○	○	○	○	○
I attend TLC meetings as a legitimate peripheral participant.	○	○	○	○	○

35. In your opinion, please indicate how strongly you agree or disagree with the statement below.

	Strongly Disagree		Agree		Strongly Agree
I am familiar with the KLT materials and able to recommend supports to teachers and TLC Teacher Leaders.	○	○	○	○	○

Demographic Background

36. Select which category best describes you.

	0-3 Years	4-6 Years	7-10 Years	11+ Years
Years of experience in education.	○	○	○	○
Years at current school.	○	○	○	○

37. What is the highest degree or level of school you completed?

○ High school diploma or GED

○ Some college credit, no degree

○ Bachelor's degree

○ Some Master's credit, no degree

○ Master's degree

○ Some Doctorate credit, no degree

○ Doctorate degree

NWEA.

Spring 2014 Keeping Learning on Track Participant Survey

Demographic Background

38. What racial category best describes you?

- ○ White
- ○ African American
- ○ American Indian or Alaska Native
- ○ Asian American or Pacific Islander
- ○ Biracial/Multiracial
- ○ Other

39. Are you of Hispanic or Latino origin or descent?

- ○ Yes
- ○ No

NWEA.

Information

40. Please provide the following information. (The information you provide will remain confidential)

First Name:

Last Name:

41. School Name:

Appendix C: Student Survey

Dear Student,

This survey is voluntary. You do not have to complete this survey, but we hope that you will. We need your help!

- Your answers will improve how well teachers teach and how much kids like you learn.
- Please do not write your name on this form. No one but you will know how you answer these questions.
- Please read carefully and circle your answer.

Example:

	Never 1	2	Sometimes 3	4	Always 5
How often do you eat vegetables?	(1)	2	3	4	5

Thank you for taking this survey!

NWEA.

The first set of questions asks about your teacher. Please tell us how often your teacher does the following:

	Never 1	2	Sometimes 3	4	Always 5
My teacher makes sure I understand what high-quality work looks like for work I do in my class.	1	2	3	4	5
When I get something wrong, my teacher clearly explains how I can do better.	1	2	3	4	5
My teacher makes sure I am clear about what we are going to learn and how we will know when we have learned it.	1	2	3	4	5
My teacher makes sure all of us respond to her/his questions in some way.	1	2	3	4	5

NWEA.

The next set of questions asks about your classroom. Please tell us how often you and the students in your class do the following:

	Never 1	2	Sometimes 3	4	Always 5
In this classroom, all students are involved in discussions.	1	2	3	4	5
In this classroom, students review each other's work to see how well we have learned something.	1	2	3	4	5
In this classroom, students give each other suggestions to improve our work.	1	2	3	4	5
In this classroom, the teacher gives us time and ways to take everyone's suggestions and my own ideas and learn from them.	1	2	3	4	5

NWEA.

The last set of questions asks about you. Please tell us how strongly you agree with the following:

	Strongly Disagree 1	2	Agree 3	4	Strongly Agree 5
I believe I can learn a lot.	1	2	3	4	5
I make sure I know what I am supposed to learn.	1	2	3	4	5
As I learn something, I think about not only what I learned, but how I learned it.	1	2	3	4	5

NWEA.