1-1-2015

Measuring Teachers' Knowledge and Use of Data and Assessments: Creating a Measure as a First Step Toward Effective Professional Development

Courtney Vidacovich
University of Denver

Measuring Teachers' Knowledge and Use of Data and Assessments: Creating a Measure as a First Step Toward Effective Professional Development

Abstract

Current teaching standards and practices are dictated, at least in part, by state- and district-mandated standardized tests. Yet, despite being surrounded by data, teachers receive only basic trainings on how to use assessments. In reality, teachers use data and assessments daily--even minute by minute--through the assessment process, which uses multiple data sources to make informed decisions on student learning and teaching practices. A measure was needed to understand how the policies and expectations from schools, districts, and states compare with actual classroom practices. The teachers Knowledge and Use of Data and Assessment (tKUDA) measure was designed to do just that.

This study sought to create and assess the validity of the tKUDA while exploring differences between respondents and relationships between factors. The tKUDA is a tool with support for reliability and validity to be used to gauge teacher practice around data and assessments through the assessment process. Reliability was assessed via Cronbach's alpha (Knowledge factor alpha = 0.95, Use factor alpha = 0.96) and using item response theory (Knowledge person separation = 2.52, reliability of person separation = 0.86; Use person separation = 1.11, reliability of person separation = 0.55). Validity was evidenced through correlations between expert interview ratings and item difficulty ($r = 0.87$), correlations between similar, known measures and the tKUDA ($r = 0.41$). Additionally, construct validity was seen through scale use and internal validity was presented using differential item function.

The tKUDA allows administration, university teacher preparation programs, and researchers to identify strengths and needs of teachers in order to create meaningful, targeted training opportunities. Differences were seen between Knowledge and Years of Teaching and between Use and Content Expertise. A moderate, positive relationship between Knowledge and Use was found with Knowledge explaining 22% of Use. Evidence for possible differences in this relationship by content are also noticed and discussed.

Document Type

Dissertation

Degree Name

Ph.D.

Department

Quantitative Research Methods

First Advisor

Kathy E. Green, Ph.D.

Second Advisor

Nicholas Cutforth

Third Advisor

Richard Kitchen

Keywords

Assessment, Classroom assessment, Data use, Measurement, Professional development, Teacher

Measuring teachers' knowledge and use of data and assessments:

Creating a measure as a first step toward effective professional development

---

A Dissertation

Presented to the Faculty of the Morgridge College of Education

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Courtney Vidacovich

November 2015

Advisor: Dr. Kathy Green

Author: Courtney Vidacovich
Title: Measuring teachers' knowledge and use of data and assessments: Creating a measure as a first step toward effective professional development
Advisor: Dr. Kathy Green
Degree Date: June 2015

ABSTRACT

Current teaching standards and practices are dictated, at least in part, by state- and district-mandated standardized tests. Yet, despite being surrounded by data, teachers receive only basic trainings on how to use assessments. In reality, teachers use data and assessments daily—even minute by minute—through the assessment process, which uses multiple data sources to make informed decisions on student learning and teaching practices. A measure was needed to understand how the policies and expectations from schools, districts, and states compare with actual classroom practices. The teachers Knowledge and Use of Data and Assessment (tKUDA) measure was designed to do just that.

This study sought to create and assess the validity of the tKUDA while exploring differences between respondents and relationships between factors. The tKUDA is a tool with support for reliability and validity to be used to gauge teacher practice around data and assessments through the assessment process. Reliability was assessed via Cronbach's alpha (Knowledge factor alpha = 0.95, Use factor alpha = 0.96) and using item response theory (Knowledge person separation = 2.52, reliability of person separation = 0.86; Use person separation = 1.11, reliability of person separation = 0.55). Validity was evidenced through correlations between expert interview ratings and item difficulty (r = 0.87), correlations between similar, known measures and the tKUDA (r =

0.41). Additionally, construct validity was seen through scale use and internal validity was presented using differential item function.

The tKUDA allows administration, university teacher preparation programs, and researchers to identify strengths and needs of teachers in order to create meaningful, targeted training opportunities. Differences were seen between Knowledge and Years of Teaching and between Use and Content Expertise. A moderate, positive relationship between Knowledge and Use was found with Knowledge explaining 22% of Use. Evidence for possible differences in this relationship by content are also noticed and discussed.

ACKNOWLEDGMENTS

Thank you to all my friends, fellow graduate students, and professors who helped make this study possible. Special thanks to my advisor turned role model, Dr. Kathy Green, who has always treated me as an expert; I cannot express my gratitude for all of the opportunities, guidance, and advice you have given me. Thank you to my mother, Donna Vidacovich, for being the gold standard of what a teacher should be and to my father, Chris Vidacovich, for his love of education, boundless support, and major effort to help with data collection. Thank you to Dustin Donovan who has given me endless encouragement and makes me believe I can accomplish anything. Finally, special thanks to Kate Bolen who is the most incredible editor; I could not have done this without you and am lucky to call you family.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

Introduction

Now in its first year of implementation, Colorado Senate Bill 10-191 requires that 50% of a teacher's evaluation be based on professional practices (content, environment, instruction, reflection, and leadership) and 50% be based on student learning data. While the professional practices are guided by a statewide rubric, the Colorado Department of Education (CDE) provides very little guidance for the student learning portion of the evaluation, other than stating that it should be based on multiple measures, including a state standardized achievement test. Since 2014 was the first year of the new evaluation procedure, the state has not fully established the meaning of the student learning requirement, and districts are individually choosing how to evaluate their teachers based on student learning data ("Determining Final Effectiveness Ratings," 2013). Colorado is not the only state implementing new laws such as these. States and school districts use assessments to evaluate students, teachers, schools, and programs, but methods for evaluating teachers with assessment data are not consistent and tend to use inappropriate measures (Cai & Lin, 2006; Kellaghan, Stufflebeam, & Credo, 2003; Madaus & O'Dwyer, 1999; Shen & Cooley, 2008). States and school districts influence student assessment by teachers through policies and standards such as these. Teachers attempt to incorporate these mandated requirements into their assessment practices, but they are only able to do so effectively when guidance is given. Currently the level of guidance provided to teachers varies from district to district. This has the potential to result in inaccurate data, especially because the state is enforcing policies for teachers regarding student learning data with little understanding of how teachers are being trained and supported to use and how they are *actually* using student data.

Assessments help teachers know where students are in their learning processes, how they have grown, where they still need to develop, and how to get them there. Various assessments show teachers and students where strengths and weaknesses lie so these can be addressed. Teachers have the opportunity to reflect on their teaching practices and modify them to benefit their students. Students can understand how they have grown as learners and what areas they need to focus on. Assessments are central to determining the level of student learning and the effectiveness of teachers, but to better ascertain their influence, we need ways to understand how assessments are viewed by teachers and used at both classroom and district levels.

Jennings (2012) states that "researchers have spent much more time analyzing test score data than investigating how teachers use data in their work" (p. 1). Teachers are held accountable for results of testing and are using data to the best of their abilities, but there are serious gaps between their knowledge, the expectations placed on them by policy or administrations, and their actual practices (Greenberg & Walsh, 2012; Young, 2006). Teachers need guidance and training to bridge these gaps. To improve teacher preparation programs and professional development, researchers need to consider teachers' perspectives and daily lives in the classroom in relationship to assessment and data analysis. We need to know where to begin, based on teachers' needs, in order for these preparation programs to be meaningful and beneficial to teachers. The existing literature on teachers using student assessments and student data analysis lacks the teacher perspective, which is needed to complete the picture. As a researcher, a *psychometrician,* and a former middle school teacher, I recognize the importance of and need for creating professional development opportunities to educate teachers on ways to use their student data more effectively and efficiently. Teachers' insights on how they are using assessment and data in the classroom can inform researchers and professional developers about what teachers need to improve classroom practices regarding assessments and data use. This study offers a way to capture teachers' voices through an instrument that measures teachers' knowledge and use of data and assessments. The instrument would allow professional developers, administration, and university programs to recognize their teachers' understandings of data and assessment, discover strengths and weaknesses, and create meaningful ways to support and guide teachers based on their specific needs.

Purpose and Research Questions

This research project sought to understand how teachers use data and assessments through insights in the literature and from the creation and administration of a measure to understand teachers' knowledge and use of data and assessments. This measure is intended to be a tool that instructional coaches, professional development personnel, and university teacher preparation programs can use to gauge teachers' knowledge of and use of data and assessments in order to identify strengths and weaknesses and create specific, meaningful learning supports for teachers. The research questions for this study were:

1. What is the measured construct?
    a. Do items factor appropriately into three distinct domains of Knowledge and Use? Is the factor structure confirmed in an independent sample?
    b. Is there adequate reliability and validity for each of these factors?
    c. Are the response scales used appropriately?
    d. Is the measure well targeted?
    e. Which items are the hardest and which are easiest for the teachers to agree with?
2. Do teachers respond differently to subscales based on demographic differences, such as gender, race/ethnicity, content expertise, and number of years teaching?
3. Is there a relationship between Knowledge and Use?
    a. Does this relationship differ by demographic differences such as number of years teaching and content expertise? (i.e., is the relationship different for teachers in different content areas?)
        - H_0 – There is no relationship between Knowledge and Use of the assessment process.

The Assessment Process

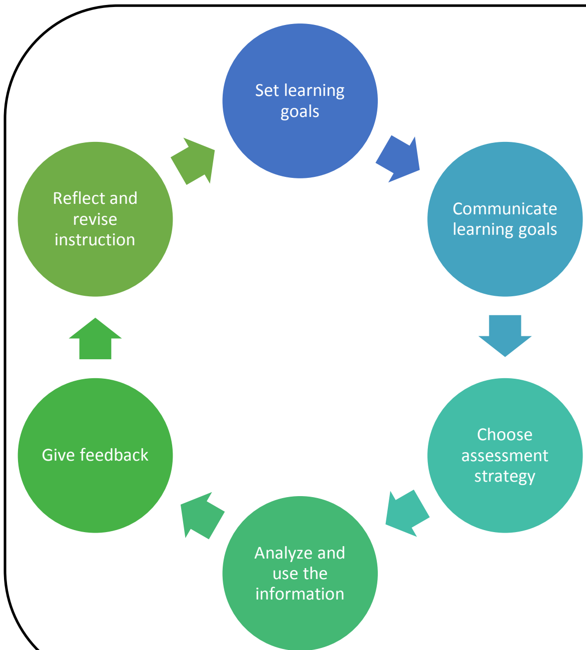
The purpose of education is to improve students' knowledge and skills, and assessments are the tools that show student learning is transpiring. The words "assessment" and "data" are often used interchangeably, but they have distinct differences. In this paper, “assessment” refers to any tools that are used to collect information about student learning, such as standardized tests, classroom tests, quizzes, observations, grades, formative assessments, student discussions, progress monitoring tools, etc. The word “data” refers to the actual information gathered from administration of these assessments, and it encompasses both quantitative data and qualitative data.

The concepts of data and assessment are further complicated because they are embedded within teaching. Assessment in education is a *process* that uses specific assessment strategies to gather evidence about student learning (National Research Council, 2001). Assessment and teaching are not separate processes; they go hand in hand (Heritage & Bailey, 2006). Heritage (2007) offers a disturbing explanation of how assessment is seen by many teachers:

> Assessment is not regarded as a source of information that can be used during instruction. Instead, it has become a tool solely for summarizing what students have learned and for ranking students and schools. In the process, the reciprocal relationship between teaching and assessment has been lost from sight. In a context in which assessment is overwhelmingly identified with the competitive evaluation of schools, teachers, and students, it is scarcely surprising that classroom teachers identify assessment as something external to their everyday practice. (p.1)

Many teachers regard assessment negatively because they view it as external to their actual work in classrooms. I witnessed this negative attitude during a pilot study in the spring of 2014 that explored how teachers defined data. During interviews, when a teacher expressed a negative opinion about data use in the classroom, often they associated assessment data exclusively with standardized testing. At this point, I explained that assessment does not only refer to testing, but rather it is a part of their teaching practice every day. The entire conversation became more positive once they understood the connection of the assessment process to their teaching.

There are various interpretations of what this process looks like in the literature (Brookhart, 2011; McMillan, 2000; Natriello, 1987; Stiggins & DuFour, 2009; Williams, 2011). Figure 1 includes the agreed upon steps of the assessment process that have been rephrased to language applicable to classroom practices. Please note that while these steps are in a specific order, they are actually fluid in real-life teaching situations (National Research Council, 2001). A teacher can go through this process over a specific period of time, like a unit of content, or this can occur several times within one class session.



1. Set specific learning goals.

2. Communicate learning goals to students in multiple ways.

3. Choose an assessment strategy that will measure the specific learning goal.

4. Analyze and use the information gathered from that assessment for student learning.

5. Give students feedback on what they need to improve and how to improve.

6. Reflect on and revise instructional practices based on the information gathered.

Figure 1. *The Assessment Process*

Set learning goals. The first step, and arguably the most important, is to set specific learning goals for the students. The National Research Council says that "educational assessment does not exist in isolation, but must be aligned with curriculum and instruction if it is to support learning" (2001, p.3). A teacher needs to be clear in the purpose of a lesson or unit so she know what she and the students are trying to accomplish. Learning goals may be called objectives, outcomes, standards, etc., but the words all mean the same thing: What should the students know and be able to do (Walvoord & Banta, 2010)? These need to be specific, carefully planned goals that are created before teaching begins (Wiggins & McTighe, 2005). McMillan (2000) stresses that good assessments enhance instruction, as long as assessment and instruction begin together.

Communicate learning goals. It is not enough to simply have a learning goal. Teachers must communicate their intentions to students so students know the purpose and have a goal to guide their own learning. In an online lecture, Williams (2013) says that teachers must clarify and share their learning intentions. He says this could occur through sharing or displaying learning objectives or through a focused question of the day (Williams, 2013). Brookhart (2011) states that "teachers should be able to articulate clear learning intentions that are congruent with both the content and depth of thinking implied by standards and curriculum goals, in such a way that they are attainable and assessable" (p.7). Most curriculum standards are not written in student-friendly language, so teachers must communicate their goals in a manner that students can clearly understand. Additionally, teachers need a repertoire of strategies for communicating learning goals to students in order to ensure that the learning goal is heard or seen and understood by all students (Brookhart, 2011; Williams, 2013).

Choose appropriate assessment type. Multiple teacher standards focus on teachers' very basic understanding of assessment types and their abilities to choose the best ones to gather the data needed. Knowing assessment types and when to use them is a major focus in teacher preparation programs regarding assessment topics (Greenberg & Walsh, 2012; Siegel & Wissehr, 2011). Picking the appropriate type of assessment is not as straightforward as it may appear. There are multiple types of assessments used in contemporary education, including observing, interviewing, questioning and discussions, curriculum-based measurements, assignment evaluations, rubric evaluations, prior knowledge assessments, self-assessments, formative assessments, summative assessments, interim assessments, comprehensive exams, performance exams, alternative exams, standardized tests, portfolios, etc. (Darling-Hammond, Bransford, & NEA, 2005; Froschauer, 2013; Stecker, Fuchs, & Fuchs, 2005). Teachers are benefited if they are able to differentiate between these types of assessment and pick appropriate ones to use in their classrooms.

Match assessment to learning goals. Teachers need to be able to explicitly match their learning objectives, like common core standards, to the chosen assessment. Stronge and Grant (2009) say that setting learning goals is all about matching student needs, standards, and prior knowledge to the assessment. Sometimes the learning goal needs a formal exam to capture student learning, but other times it might be gathered through listening to student presentations, a classroom discussion, or another method.

Shen and Cooley (2008) emphasize matching assessments to the curriculum and instruction. Stiggins and DuFour (2009) say learning targets that are translated into assessments will yield more accurate results. Shea, Murray, and Harlin (2005) strongly believe that assessments must be based on a set of accepted student learning standards. Teachers should determine learning goals early on in curriculum preparation (Walvoord & Anderson, 2009; Walvoord & Banta, 2010), and Wiggins and McTighe (2005) even assert that assessments should also be designed before instruction begins.

Analyze and use the information gathered. There are multiple ways that teachers can break down and explore assessment data, from simply evaluating work in order to determine if a student “got it” or “didn’t get it” to disaggregating data by content strands. There are also a multitude of ways that teachers use the information discovered from their assessments like differentiating instruction, reteaching content, and identifying students’ strengths and needs. These are further discussed below.

Give students feedback. One of the essential elements of the assessment process is giving students productive feedback (Heritage, 2007; Natriello, 1987). Brookhart (2011) says that all teachers should have the ability to provide students with effective and useful feedback. The feedback must be more than a grade; it should be tied to the learning goal and give guidance to help students understand what they are able to do and what they still need to do (Brookhart, Moss, & Long, 2008). Williams (2011) agrees saying that feedback should include what needs to improve and specific activities the student needs to do in order to improve. He stresses that “for assessment to support learning, it must provide guidance about the next steps in instruction and must be provided in a way that encourages the learner to direct energy towards growth" (p.7).

Reflect and revise instructional practices. Teachers may not recognize they are using data to reflect on their teaching practice and revise instruction because this use of data can be so intuitive for them. Dewey (1933) says reflection is more than just "thinking about it;" it is a methodical, rigorous way of thinking. Bruster and Peterson (2013) state:

> Reflection must be an embedded, intentionally infused component of the curriculum, involving both interaction with peers and feedback from the teacher. . . . Though developing reflective practice is a complex and potentially convoluted process, it is a means by which continued and limitless professional growth can occur. (p. 83)

Reflection is a beneficial and innate skill for most teachers. Reflective teachers look for solutions when presented with a problem; can critique solutions to decide on the best option; create and test their assumptions of students and of learning; have learner-centered, reflective classroom environments; value criticism; are very self-aware and purposeful in curricular decisions; make decisions for the future; and are very open to new professional development opportunities (Bruster & Peterson, 2013; Tok & Dolapçioglu, 2013). Shepherd, Davidson, and Bowman (2011) discovered that the majority of teachers in their study used data to reflect upon or evaluate their teaching, asking themselves questions like "If my students didn't make gains, then what am I doing that needs to change?" This is the type of reflection that leads to revisions in teaching practices; this is how actions based on data begin.

Challenges.

Teachers engage in the assessment process daily, and many do not recognize it. The assessment process becomes a routine of teaching. The cycle is never over; new understandings discovered from the assessment results lead to new goals. The challenges of this process and variations seen in teaching styles come from differing teacher knowledge of data and assessments, multiple types of assessment strategies, and multiple ways to use the information in practice.

Differing Teacher Knowledge

Teachers' knowledge of data and assessment varies drastically across K–12 classroom teachers of all content expertise. The knowledge of data and assessments that a high school math teacher has will look very different that of an elementary art teacher, but they both still use assessments. The math teacher may know more about quantitative data, but the art teacher has more knowledge of qualitative data. In this case, one teacher's knowledge of a type of data is necessarily better than the other, but this differing knowledge makes it difficult to compare and discuss *all* teachers. To begin to address this challenge I tried to understand what teachers should know about data and assessments.

There is a consensus among educational researchers, teacher preparation programs, and teacher standards about the information teachers should have about assessment and data use. Greenberg and Walsh (2012) identify three domains of assessment knowledge to include in teacher preparation programs: assessment literacy, analytical skills, and instructional decision making. While other articles recommend or imply necessary student assessment knowledge, these recommendations all fit within one of Greenberg and Walsh's three domains. Assessment literacy refers to a teacher's basic understanding of assessment types and his or her ability to develop classroom assessments. Analytical skills refer to a teacher's understanding of how data are used and his or her ability to dissect, describe, and display data. Instructional decision making refers to a teacher's ability to guide instruction with evidence of learning from student data. Greenberg and Walsh evaluated teacher preparation programs and created a rubric to better demonstrate different levels of knowledge in each domain (Table 1). Ideally, all teachers should have the understandings from category 4, but they found that most preparation programs only showed evidence from categories 2–3 for assessment literacy, 1–2 for analytical skills, and 3–4 for instructional decision making (Greenberg & Walsh, 2012).

Table 1
Greenberg and Walsh's (2012) Rubric for Knowledge Domains

Assessment Literacy				
0	1	2	3	4
There is no or almost no instruction or practice on the various types of assessment	Instruction on the various types of assessment is very limited and there is no or almost no practice	The scope of instruction on the various types of assessment is EITHER not comprehensive and practice is limited OR instruction is comprehensive but practice is very limited	The scope of instruction on various types of assessment is comprehensive and there is adequate practice	The scope of instruction on various types of assessment is comprehensive, including concepts of validity, and reliability, and there is adequate practice

Analytical Skills				
0	1	2	3	4
There is no or almost no practice or instruction preparing teachers to analyze data from assessments	Instruction preparing teachers to analyze data from assessments is very limited, and there is very limited practice	The scope of instruction on analyzing data from assessments is EITHER not comprehensive, but practice includes field-based practice and/or presentation of quantitative data and graphs OR instruction is comprehensive but practice is limited	The scope of instruction to prepare teachers to analyze data from assessments is comprehensive and practice includes field-based practice and/or presentation of quantitative data and graphs	The scope of instruction on analyzing data from assessments is comprehensive and practice includes field-based practice AND presentation of quantitative data and graphs

Instructional Decision Making				
0	1	2	3	4
There is no or very limited instruction or practice that prepares teachers to use assessment data to drive instruction	There is limited instruction or practice that prepares teachers to use assessment data to drive instruction	Instruction or practice that prepares teachers to use assessment data to drive instruction is evident but only in one or two subject areas	Instruction or practice that prepares teachers to use assessment data to drive instruction is evident but not in all subject areas	Instruction or practice that prepares teachers to use assessment data to drive instruction is evident in all subject areas

Assessment Literacy. Teachers should know about different types of assessments, understand concepts of validity and reliability, and be able to develop classroom assessments (Greenberg & Walsh, 2012). Stiggins (1991) says, “Those who are assessment literate have a basic understanding of the meaning of high- and low-quality assessment and are able to apply that knowledge to various measures of student achievement" (p. 535). Earl and Katz (2006) describe this as data literacy, saying it is a process: decide what you need to know, collect the appropriate data, find ways to connect various data sources, ensure data are worth considering, be aware of the limitations of the data, and think about what the results mean. Walvoord and Anderson (2009) include these ideas in their process for assessing students' work: the first two steps are to identify what information is needed and then measure student work against learning goals. Both pairs of researchers agree that teachers need to know different types of assessments to capture the types of data needed to judge student learning, make sure the data are trustworthy, and be able to create the needed assessment.

Assessment literacy is the first step in understanding data and assessments that all teachers need, but Knowledge and Use are very different constructs. Siegel and Wissehr (2011) conducted a study to explore pre-service teachers' assessment literacy skills through detailed document analysis. Their findings noted that teacher candidates recognized different assessment tools, understood their importance, and knew the power of using multiple assessments to evaluate learning, but this did not transfer into their student teaching experiences.

Analytical Skills. Teachers should be able to use analytical skills to dissect, describe, and display assessment data. They should also understand concepts of error, types of scores, growth versus performance, and triangulation of student data (Greenberg & Walsh, 2012). For pre-service teachers, this is the intersection between pedagogy classes and an introductory statistics course. This can initially seem intimidating for an educator who is less familiar with statistical analysis. However, teachers need to be taught to have at least a basic understand of data analysis instead of relying on outside sources or being intimidated by the terminology used by data analysts (Taylor, 2009). Even if teachers are not comfortable analyzing their own data, they should know the words and concepts behind the data and student scores that are given to them.

Hoover and Abrams (2013) conducted a national survey and found that teachers have access to a lot of data and are skilled at data collection, but they are only analyzing data at an aggregate level with central tendencies, typically just using mean scores. Few teachers are going deeper in analyzing their data. This could be due to teachers having limited access to useful analysis methods and programs, as indicated by much of the literature. Marzano (2003) says teachers and schools must have a system in place to analyze the data; some sort of explanatory model is needed to help make sense of the numbers or observations. Shen and Cooley (2008) state "Many districts do not possess the technological infrastructure to analyze data in a form for efficient and effective use by teachers." Wong and Lam (2007) provide a guide on how teachers can disaggregate student data using Excel, but it is a very complex process. McDonald (2002) argues that teachers do not have adequate professional development to help them make the most of their student data. Marsh, McCombs, and Martorell (2009) found a significant association between teachers who were given support with data analysis and those teachers' perceptions of themselves and student achievement scores. Walsh (2003) conducted a study of several schools that are closing the achievement gap and noticed these teachers were more likely to participate in specific professional development on analyzing student data, especially for their lower-achieving students.

There is only one standard in Colorado for teachers that matches analytical skills, and it relates to using multiple data sources to triangulate student learning ("Rubric for evaluating Colorado teachers," 2013). Triangulation can be completed with quantitative data, qualitative data, or a mix of the two. Teachers seem effective at not relying on one data source, but instead they use evidence from multiple assessments and observations. It is interesting to note that triangulation does not focus heavily on quantitative skills, which may be why teachers are more inclined to use this analytical skill as part of their assessment practices.

Analytical skills should bridge the gap between assessment literacy and knowing how to do something with the results. Unfortunately, based on the literature and the teacher standards, this does not appear to be the case for many teachers. Little (2012) points out the dilemma of teachers shying away from discussing data, instead discussing instructional factors associated with student learning; also, teachers tend to focus on processes more than actual meanings behind the data. Hoover and Abram (2013) also imply that teachers are likely to skip the analysis step. This concern needs to be addressed more thoroughly in pre-service education and professional development to aid teachers.

Instructional Decision Making. Instructional decision making refers to a teacher's ability to guide instruction using evidence from student data. Teachers should practice data-driven decision making using formative assessments, backward design lesson planning, and an understanding of the instructional implications of data, and they should know how to identify student misunderstandings (Greenberg & Walsh, 2012). The literature that focuses on these ideas is expansive, and only a few conclusions from the literature are highlighted here. Teacher standards are also focused heavily in this category (Association of Childhood Education International, 2007; National Council for the Social Studies, 2002; National Council of Teachers of English, 2012; National Council of Teachers of Math, 2003; National Science Teacher Association, 2012; Rubric for evaluating Colorado teachers, 2013). Instructional decision making appears to be central in teacher preparation programs and in a teacher's daily life.

In 1983, the National Education Association of the United States (NEA) said, "The NEA holds that testing and assessment should be conducted frequently, be comprehensive in nature, and serve educational purposes. Testing and assessment should be carried out to diagnose student weaknesses and strengths." This is even more true 30 years later. Instructional decision making is the link between discovering problems and creating solutions based on data. There is a recent push in education that is focused on data-driven decision making. McLeod (2009) identifies five main elements for accomplishing this: create a way to get good baseline data to know where students are beginning; set clear, measureable goals; conduct frequent formative assessments; discuss data in professional learning communities; and create instructional interventions based on needs seen in the data.

Formative assessments, assessments for learning, are a significant part of instructional decision making because these tools and the resulting data are used to monitor progress (Shea et al., 2005). Formative assessments are a crucial part of goal setting and monitoring progress (Strong & Grant, 2009), and this kind of assessing should be implemented daily and used to drive instruction (Shea et al., 2005). Shen and Cooley (2008) think that "an overemphasis on achievement data based on standardized tests does not provide a clear student learning profile and has limited implications for curriculum and instruction" (p. 321). Formative assessments are classroom-specific and focused on current practices happening in the moment for students and teachers. This makes formative assessments more useful for decision making than standardized tests, which are removed from the work of individual classrooms.

Differing Assessment Strategies

The assessment process is also confounded due to the multitude of assessment strategies that teachers can choose to use. Again, different teachers will rely on different types of assessment strategies; assessments in a math class look different that those in the art class, and neither approach is wrong. All assessments are designed to evaluate student learning for some intended purpose. There are endless ways that a teacher can assess her students. In order to try to explain these numerous strategies in a concise manner, I will consider assessment types as two categories: formal and informal.

Formal Assessments. Formal assessments are documented, performance assessments. They typically occur at the end of a learning period and are typically used to demonstrate student achievement. Standardized tests are formal assessments, but this category can also include non-standardized, classroom exams. There are different levels of formal assessments that can be seen and used by classroom teachers. These include state-level exams, like the yearly state test; district-level exams, like interim and benchmark tests; school-level exams, like common assessments or end of year exams; and classroom-level exams, like unit tests, projects, presentations, quizzes, etc.

Each type of formal assessment is valuable to a teacher if considered in the context of the intent of the exam. Shepard (1989) says classroom assessments are less reliable statistically than formal assessments, but can gather data about individual students over a school year in a much more accurate way than an annual standardized test. The research recommends great caution about how teachers should be using the data gathered from summative assessments created by outside sources, like state and national standardized exams. Standardized tests play a part in education, but should not be used as the primary indicator of student learning (National Research Council, 1999). Educators can devise ways to raise standardized test scores, but this is not the same thing as improving student learning (Shea et al., 2005). Standardized tests are created to monitor broad policy questions, evaluate educational programs, to analyze trends over time, or for accountability (Darling-Hammond et al., 2005). As indirect measures of student learning, these tests are not appropriate for accurately measuring learning (Marzano, 2003). An overemphasis on achievement data that is based on standardized tests does not provide a clear student learning profile and has limited implications for curriculum and instruction (Shen & Cooley, 2008).

So what can teachers use these data for? Darling-Hammond et al. (2005) have several recommendations. State tests should be based on curriculum frameworks and conceptual goals that are the same as the ones teachers are using. If that is truly the case, these tests can be used by teachers to evaluate their own overall curriculum and instruction for strengths and weaknesses. Another way is to match similar cognitive skills from the state exam to teachers' own unit tests by using similar tasks, knowledge level of items, standards to evaluate student work, etc. If teachers have the ability to disaggregate their own data, they can compare previous standardized exam scores to current data and potentially predict outcomes on future standardized tests. Means, Gallagher, and Padilla (2007) found that teachers could use these larger exams to monitor student progress, inform their curriculum planning, and refine their instruction based on the scores. Monpas-Huber (2010) noticed that teachers use this type of data more frequently when it is easily accessible, can be used effectively, and is used to improve school-level performance on the state assessment. Even though teachers are able to use this data source in specific ways that are potentially beneficial to student learning, they should use extreme caution to ensure that the exam results are appropriate for what they are being used for. These exams are more valuable for accountability than for instructional improvement.

Informal Assessments. Informal assessments are typically not documented and are used to gauge student learning. These typically occur during the learning period and are intended to evaluate students' understanding of the content. These assessments only occur at the classroom-level and are synonymous with formative assessments. Examples of this type of assessment include exit slips, questioning, fingers 1–5, white boards, classroom or peer discussions, observations, etc. Heritage (2007) identifies three broad categories of informal assessments: on-the-fly assessments, planned-for interactions, and curriculum-embedded assessments. These assessments tend to be underrated by administration and policymakers, but these are the ones teacher tend to rely on the most (Shen & Cooley, 2008).

Differing Classroom Practices

The assessment process also varies due to how teachers are actually using the information gathered from their assessment strategies. The way the math teacher and art teacher use data is going to look very different in practice, but should be similar in pedagogy. For example, they should both be differentiating their instruction based on data, but how this manifests in each classroom will be different.

Teachers excel at data collection, but opportunities to use data are varied (Earl & Katz, 2006; Hoover & Abrams, 2013). The literature is vast regarding assessment in education, but it is quite limited regarding actual teacher practices about how they use student data. Most published articles describe current conditions, show exemplar situations, or recommend how things *should* be done, but few explore what teachers *actually* do. This section summarizes literature recommendations and teacher assessment standards indicating what teachers should know and what they should be able to do. This does not mean that teachers actually know or do these things.

The literature on how teachers should use assessments and data is varied and no framework was found. In order to create a working understanding of teachers' use of data, I compiled a list of recommendations from the literature and teacher standards and categorized them into distinct concepts. The search was extensive, but approximately 60 books and articles fit the content parameters based on teacher use of data. In addition to the literature, teacher standards were vital in producing this list of recommendations on assessment and data use. If a standard is in place, it can be assumed a teacher should be able to do it. Included standards are sourced from the Association of Childhood Education International (2007), the Colorado Teacher Quality Standards (2013), the National Council for the Social Studies (2002), the National Council of Teachers of English (2012), the National Council of Teachers of Math (2003), and the National Science Teacher Association (2012). This study originates in the state of Colorado, which passed new teacher evaluation procedures in 2013, so the Colorado Student Learning Evaluation Procedures (2012) and the Colorado Performance Practices Rubric (2012) were also considered. After categorizing the information, there are ten aspects that the literature, standards, and Colorado evaluation agree teachers should be able to do using assessments and data:

1. Drive or inform instruction
2. Set learning goals
3. Reteach or review content
4. Differentiate instruction
5. Evaluate student learning

6. Reflect on and revise instruction based on data
7. Give specific feedback on student learning
8. Identify student strengths and weaknesses
9. Triangulate student learning
10. Disaggregate student data

Details on how these aspects are used in the classroom are explained more below, except for "Set learning goals," "Reflect on and revise instruction based on data," and "Give specific feedback of student learning," which were explained above in the Assessment Process section.

Drive or inform instruction. The concept of teachers using data to drive instruction is an important use of data that truly focuses on matching the data to the instruction. "To realize educational excellence we must go beyond analyzing student problems to developing solutions" (NEA, 1983). Walvoord and Banta (2010) state "The end of assessment is action" (p. 2). Hamilton, Halverson, Jackson, Mandinach, Supovitz, and Wayman (2009) believe teaching should be a continuous cycle of collecting data on students' learning, which leads to interpreting data to develop a hypothesis on how to improve learning, which leads to modifying instruction to test this hypothesis, which returns to collecting data. McLeod (2009) discusses a similar concept: create pre-assessments to establish good baseline data, set clear and measureable goals, conduct formative assessments throughout the learning period, analyze data and discuss with others, and then create instructional interventions based on needs seen in data. This is where formative assessments are vital by helping to monitor progress throughout the learning period. Schildkamp and Kuiper (2010) noted that teachers mainly use classroom-level data for making instructional decisions at the classroom level; summative data are not as information-rich for this purpose. Hamilton et al. (2009) describe several possible ways that teachers can modify instructional practices based on findings from student data, including prioritizing more time for particular students, reordering curriculum paths, giving individual students specific instruction, identifying instructional interventions that can help students progress, gauging effectiveness of classroom lessons, refining instructional methods, and attempting new ways to give instruction on topics. Whatever the method used, giving assessments and examining data is not the end goal; modifying instruction based on student need is the main goal. Teachers need to be able to use data to drive their instruction.

Reteach or review content. Teachers reteach, revisit, or review content all the time. This is a practical application for assessment data (Shepherd, 2011) since the decision to reteach does not occur on a whim. Teachers reteach or revisit curriculum as needed by their students. The decision to stop and reteach a concept can come from many data sources; a teacher may give a quiz, assign homework, or many teachers can simply read students faces and frustration levels. Teachers may plan a review session before a formal test, but the specific content they choose to include in the review session is based on some information or concern that students may need more time to reflect on an idea. Frohbieter, Greenwald, Stecher, and Schwartz (2011) state that reteaching and reviewing content is a key way that teachers use data.

Differentiate instruction. Another practical way that teachers use data is to differentiate their instructional practices. Differentiation is an instructional technique that allows teachers to reach students at various levels instead of teaching to just one type of student (Good, 2007). With 30 students in a classroom it is challenging for a teacher to teach in a manner that reaches students in the middle while supporting the lower achieving students and pushing the higher achieving students; but that is exactly what the teacher is asked to do. There are various ways that teachers differentiate including creating student groups, individual tutoring, creating different assignment levels, etc. (Tomlinson, 2000). The way to effectively accomplish differentiation is through use of some type of data that identifies levels of student achievement. Shepherd et al. (2011) list grouping students, individual tutoring, and class placement as effective ways teachers use data. Once a teacher has accurate information on student achievement, then she has a starting point for a student and can aid in that student's growth.

Evaluate student learning. Teachers evaluate student learning daily, and this aspect specifically refers to a basic ability to analyze student work for learning. It indicates that a teacher is able to understand whether or not a student "gets it." This could be as simple as assigning a grade. Evaluating student work to check for learning is so inherent in education that a teacher may not consider it to be using data. Little (2003) noted the concept of "looking at student work" to be a relevant data analysis practice, especially within professional learning communities. Evaluating student work does not need to be from formal data; it can include classroom work, homework, student discussions, presentations, various assignments, etc. (Nelson, Slavit, & Deuel, 2012).

Shea et al. (2005) recommends that this evaluation take into consideration student thinking, student understanding, student ability, and correctness.

Identify strengths and needs. Teachers may seem to inherently know their students' strengths and needs, but their intuition is based on some data source. It may appear at first that identifying an individual student's strengths and needs is an easy way to use data, but being able to pin point these is a practiced skill. For example, there is a difference in saying that a student's strength is math and his weakness is writing, versus a teacher's ability to use assessment data to point out that he is skilled at number sense but lacks the vocabulary needed to write detailed stories. Shepard (1979) states that the purpose of many assessments is to identify student strengths and weaknesses, but we should be careful to rely too heavily on one piece of data when doing so.

Triangulate student learning. By using multiple types of assessments and comparing them, teachers can triangulate student learning. In 1983, the National Education Association said, "Decisions about students should be based on both pencil and paper tests and a broad range of assessment methods available to teachers" (p. 1), and this is still true today (McMillan, 2000; Walvoord & Banta, 2010). Student achievement data not connected with other types of data will not facilitate student learning (Shen & Cooley, 2008). Birrell and Kee (1996) say using "multiple measures of assessment, including classroom observation and anecdotal records, samples of children's work, and children's self-assessment, can be meaningful alternatives or additions to standardized testing" (p. 286). Teachers use a variety of forms of data (Datnow, Park, & Kennedy-Lewis, 2012), but there does not appear to be a "right way" to intersect these data, just an importance of using multiple measures to gain a more complete picture of student learning.

Disaggregate student data. When teachers break information down by content standard, gender, race/ethnicity, socioeconomic status, etc. to compare achievement and growth across groups, they are disaggregating student data. It is important to note that there are no teacher standards that match this component of data use, but it is included in the framework because of its presence in school discussions and the recommendations found in the research literature. Managing and disaggregating data in this manner is not adequately taught in teacher preparation programs (Greenberg & Walsh, 2012). Teachers do not have appropriate software available to help with analysis beyond Excel, and it is unlikely they have been taught how to use Excel for these purposes (Wong & Lam, 2007). Even so, researchers, policy-makers, and school administrators seem surprised to find that teachers are not disaggregating their data. If teachers were able to perform more advanced data analyses, it could be possible to match their classroom assessments with standardized exams in a much more meaningful way and better predict where students fall on the spectrum of unsatisfactory to advanced. If the expectation of using data in this manner truly exists, disaggregating data appears to be a major professional development need.

Differing Intentions: Formative vs. Summative

Additionally, the way these data and assessments are used in practice will differ depending on the intended purpose: formative or summative. Assessments that are formative in nature are used to improve learning, gauge student understandings, and move learning forward. Summative assessments are ones that are used to *prove* learning occurred; these are typically the more traditional tests. Even though the intentions of improving learning versus proving learning are associated with specific assessment types, any assessment strategy can be used (Hoover & Abrams, 2013; Shepherd, 2012; William, 2013). For example, a quiz is typically used to prove student learning, but if it is not graded or students have multiple chances, then it is used to improve their learning. Another example is if a teacher gives an exit ticket that is typically used to gauge learning, but she grades it, then it becomes a tool to prove learning. The intention of the teacher, how the data are being used, determines if it is a formative or summative assessment.

Formative Assessments. Stiggins (2002) identifies a critical need for assessments that are for learning, ones that assess if learning is happening. When Shea et al. (2005) say that assessments occur every day, throughout the day, they are referring to various types of formative assessments. Examples of formative assessments include, but are not limited to, questioning, interviewing, student work, curriculum-based measures, quizzes, exit slips, or any other tool to gauge where a student is on a topic (Chen, Crockett, Namikawa, Zilimu, & Lee, 2012; Stecker et al., 2005). Quizzes are often considered summative assessments because they can be used to demonstrate learning, but there is a newer focus on teachers using summative assessments in a formative manner (Hoover & Abrams, 2013); it depends on the purpose of use. Formative assessments should be designed so that they aid learning, not necessarily show learning (Torrance & Pryor, 2001).

All methods of formative assessments have similarities in their design. Based on an extensive literature search, Supovitz (2012) identified three key types of information to include when designing formative assessments: information on students' development, information about students' thinking processes, and information about misconceptions. Shea et al. (2005) also stress the importance of gaining evidence of students' thinking. Hunt and Pellegrino (2002) state that formative assessments need to take place during learning and not at the end of a learning period, which is how these assessments guide instruction. They also point out that these should be student-focused assessments that allow students to show what is known, instead of being teacher-focused to judge what was learned (Hunt & Pellegrino, 2002). Stronge and Grant (2009) think formative assessments are a crucial part of setting goals for students.

So how does a teacher actually create formative assessments? Darling-Hammond et al. (2005) have a three-part model for creating formative assessment: set learning goals to identify where you are trying to go, use formative assessments to identify where you are now, and then use learning activities to decide how to achieve the learning goals. This is very similar to Wiggins and McTighe's (2005) "backwards design" curriculum planning scheme of setting learning goals first, then choosing how you will assess these goals through formative or summative assessments, and only then choosing appropriate learning activities. Either method shows these are created before learning and conducted throughout a learning period.

The purpose of formative assessments is to be used to improve learning and thus drive instruction (Shen & Cooley, 2008; Shepard, 2009). In a seminal study, Black and William (1998) conducted a meta-analysis and noted effect sizes regarding formative assessment practices to be between 0.4 and 0.7, which are quite large for education. Others have also noticed significant increases in student scores when teachers use formative assessments to drive instruction (Barnett, 2011; Kline, 2013; Wilson, 2009). These studies found that using formative assessments correctly impacted the students' learning.

Formative assessments should be used to improve learning by gauging students' understanding at various points in time throughout the learning period. These assessments, in any form, are used to monitor student progress (Stronge & Grant, 2009), and they should be student-centered and allow students to monitor themselves, see progress in their learning, and understand areas for improvement (Hunt & Pellegrino, 2002). If used in this manner, formative assessments can also promote conceptual changes in students that go beyond classroom learning to self-efficacy and metacognitive skills (Yin, Shavelson, Ayala, Ruiz-Primo, Brandon, Furtak, Tomita, & Young, 2008).

Summative assessments. Summative assessments are assessments of learning (Stiggins, 2002) and document student achievement at a specific point in time (Darling-Hammond et al., 2005). Many in education associate summative assessments with standardized exams, but summative assessments can also be teacher-made multiple-choice tests, performance exams, essays, presentations, final projects, etc. Herman

(1992) says, “Good assessment is built on theories of learning and cognition, and it assesses the skills students will need for future success” (p. 75).

There is agreement on how summative assessments should be created in a classroom setting (Darling-Hammond et al., 2005; Walvoord & Anderson, 2009; Wiggins & McTighe, 2005). Summative assessments should be created before learning happens, and they should be based on specific, set learning goals. Providing clear goals to students helps make the goals more likely to be achieved; it helps students understand what is important and should be focused on. Teachers should ask themselves, “By the end of this course my students should know . . . and be able to do . . .” and use this to decide on the appropriate test. Summative testing can help students focus on and engage with the content in a different, deeper way than during learning, but these assessments need to be created in a manner that is authentic to the learning that occurred. Teachers should identify key vocabulary and concepts and use these to create test items that tie back into learning goals. It is also very important to consider the level of student thinking involved in the test creation process. Teachers should consider Bloom’s taxonomy to gauge if test items are measuring knowledge, comprehension, analysis, evaluation, etc. (Darling-Hammond et al., 2005; Walvoord & Anderson, 2009; Wiggins & McTighe, 2005).

There is agreement across the literature that when mixed with other forms of assessment, summative assessments help gain a complete understanding of what the student knows and can do (Shea et al., 2005). Otherwise the literature and teacher standards are split into what will be considered two levels of competency in use: basic and advanced.

Most teachers appear to have a basic understanding of how to use summative assessments. These teachers use summative assessments to reflect on their teaching practices, consider revisions for the next year or unit, and decide if it will be necessary to reteach concepts (Darling-Hammond et al., 2005). Teachers also use this for accountability purposes and to help show that students are learning (Shen & Cooley, 2008). These teachers consider patterns across the data, and many use central tendencies, but they typically are not looking deeper into the analysis.

Advanced use of summative assessments happens when teachers complete basic use and also look deeper into the data. These teachers are disaggregating their student data based on learning goals, standards, classes, student demographics, comparisons across teachers, etc. (Hoover & Abrams, 2013). They may still rely mostly on central tendencies and not more advanced hypothesis tests, but these teachers are at least trying to go deeper into data analysis.

Hoover and Abrams (2013) conducted a survey of over 650 elementary and secondary teachers to determine how they used summative assessments and data. Of those surveyed, 85% of teachers reported giving internal, summative tests or quizzes on a weekly basis, but only about 35% analyzed these with central tendencies. A majority (80%) of teachers gave quarterly benchmark exams, but only a third reported disaggregating these data by content standards or student demographics. This demonstrates there is a large difference between giving exams and using the results of exams. Additionally, of the teachers who gave summative exams, over 90% said they use the results to make changes to their instruction. Hoover and Abrams state,

> Although this large percentage suggests that teachers are considering summative assessments, evaluating and changing their instructional practices on the basis of assessment information, the low number of teachers reporting the use of more powerful forms of analysis . . . calls into question the depth of the information used to inform instruction. (p. 228)

There are other studies that have noted the same phenomenon (Earl & Katz, 2006; Greenberg & Walsh, 2012; Wong & Lam, 2007). At the very least, Brookhart (2011) says teachers should be able to accurately interpret test results in order to make decisions about students, classrooms, schools, and districts.

Researcher Voice

Before I dive deeper into the problem and purpose for this study, I feel it is important to address my voice and bias as the researcher. As a former middle school math and science teacher, I am biased toward the teacher perspective. I agree wholeheartedly with Jennings's (2012) statement that "researchers have spent much more time analyzing test score data than investigating how teachers use data in their work" (p.1). I feel that the top down approach has not been working as intended and this disconnect is seen in the literature explained above. Teachers are being held accountable and are using data to the best of their ability, but they need more guidance and trainings to be effective. Teachers are not stupid or lazy; we just have a lot on our plates. It is hard to implement change all the time. Teacher knowledge and teacher practice are different; if the actual practices do not match the knowledge, it is most likely because other aspects of the teaching life get in the way. I strongly believe that for programs to be effective, teachers need to know where to start. I cannot even count the number of teacher trainings I attended that repeated the same learning and never moved us forward.

Worse still were the trainings that changed every year so that we had no consistency or chance to build on pre-existing knowledge. The literature has very little teacher-voice included on this topic, and this is needed to complete the picture. My hope is to combine my knowledge as a researcher, a *psychometrician*, and a former classroom teacher to fill a gap and connect themes in the literature that will lead to many needed next steps in creating meaningful professional development opportunities for teachers to use data more effectively and efficiently.

Helping Teachers to Use Data and Assessments: Create a Measure

In order to create effective, meaningful trainings in university programs and professional development, researchers, universities, and district personnel need to understand assessment and data analysis of student learning from a teacher's perspective. Once we understand how teachers are currently using assessments and data in their classrooms, then we can create ways to aid them in using those tools in the effective, meaningful manner that the National Education Association was calling for in 1983. There are three main problems that need to be addressed when considering matching teachers' knowledge and use of data and teacher evaluations. The first is that there is too little teacher perspective present in the literature to indicate what they know and how they can use data and assessments. Second is the struggle of connecting teacher evaluations to classroom practices and having a consistent and comparable means of measuring them. Finally, we need to have the ability to consistently measure teachers' use of data in the classroom, but no appropriate measure exists. These problems could be addressed with the creation of a measure that addresses knowledge, use, about data and assessment from a teacher's point of view for the following reasons.

First, with new accountability laws implemented in the last 30 years, teachers and students have become used to various types of assessments; these are a part of life in the education world. Studies have been conducted on administration views of data, as well as students' views of data, and while there are also a small number of studies that consider how teachers perceive data use, more are needed. Because teachers are the biggest single influence on student achievement (Nye, Konstantopoulos, & Hedges, 2004), their voices and perspectives should be the most influential in policy decisions, yet their point of view is consistently overlooked or ignored.

Second, evaluations are significant for a teacher's career. A good evaluation can have a strong impact on improving a teacher's practice (Milanowski, 2004). Teacher evaluations should be helpful and collaborative (Taylor & Tyler, 2012), but they do not always capture what really happens in the classroom. A 30-minute observation is likely to be insufficient for demonstrating how a teacher is incorporating student data into his or her teaching practice, and teachers are not always able to articulate these practices in evaluation meetings. Teachers need help understanding and explaining what they naturally do in their classrooms, and administrators need a better way to measure these practices.

Finally, some type of measure is needed to capture these ideas. To date, only three extant measures were identified, and none are comprehensive or appropriate to explain the actual classroom setting and teacher's daily lives. The Assessment Literacy Inventory was created by Mertler and Campbell (2005) to try to capture teachers' assessment understandings. This scale is reported to have a reliability estimate of 0.74 (Merlter & Campbell, 2005), which is adequate for research purposes but not for individual diagnosis. Perry (2013) created a modified version of this scale called the Classroom Assessment Literacy Inventory and this is seen in a few other studies as well. He found a reliability of 0.54 for current teachers and 0.74 for pre-service teachers. In addition to reliability concerns, these scales are measuring only a piece of the picture; there is more to assessment than knowledge and basic understanding. Lysaght and O'Leary (2013) created an observational instrument to measure how teachers use formative assessments. While this is a much more reliable measure with alphas ranging from 0.82–0.92, the problem again is that it is only capturing a piece of the overall construct. Finally, Cavalluzzo, Geraghty, Steele, Holian, Jenkins, Alexander, and Yamasaki (2014) created a survey called Using Data to Inform Decisions designed to understand how math teachers use data to inform their instruction. This survey has three parts: knowledge and skills scale, attitudes and beliefs scale, and the data use scale. The internal consistency reliability of all three scales together was 0.67, but the internal consistency reliability of the data use scale alone was $\alpha = 0.87$. Cavalluzzo et al.'s survey had higher estimates of reliability, but captures only a section of teachers. A more comprehensive picture is needed to connect the literature and aid teachers in using data and assessments more effectively and efficiently. Therefore the goal of this study was to create a measure with evidence of adequate reliability and validity that captures a more comprehensive view of teachers' classroom practices.

CHAPTER 2: METHOD

The purpose of this measure development study was two-fold: to create an instrument that can measure teachers' Knowledge and Use of data and assessment and then to explore differences between respondents and relationships between factors. There were four main research questions for this study:

1. What is the measured construct?
    a. Do items factor appropriately into distinct domains of Knowledge and Use? Is the factor structure confirmed in an independent sample?
    b. Is there adequate reliability and validity for each of these factors?
    c. Are the response scales used appropriately?
    d. Is the measure well targeted?
    e. Which items are the hardest and which are easiest for the teachers to agree with?
2. Do teachers respond differently to subscales based on demographic differences, such as gender, race/ethnicity, content expertise, and number of years teaching?
3. Is there a relationship between knowledge and use?
    a. Does this relationship differ by demographic differences such as number of years teaching and content expertise? (i.e., is the relationship different for teachers in different content areas?)

- H_0 – There is no relationship between Knowledge and Use of the assessment process.

4. What are the practical applications of data and assessments that teachers rely on and what data sources are they using?

This was a measure development study that can be described as an exploratory mixed methods design (Creswell & Plano-Clark, 2011). The qualitative data were gathered first though focus groups, interviews with experts, and cognitive interviews to create and validate survey items. The quantitative data were collected through a small pilot sample to verify and modify items as needed and then through a final field administration. DeVellis (2003) describes this process in four stages: planning, construction, quantitative evaluation, and validation. A design diagram is seen in Figure 2.

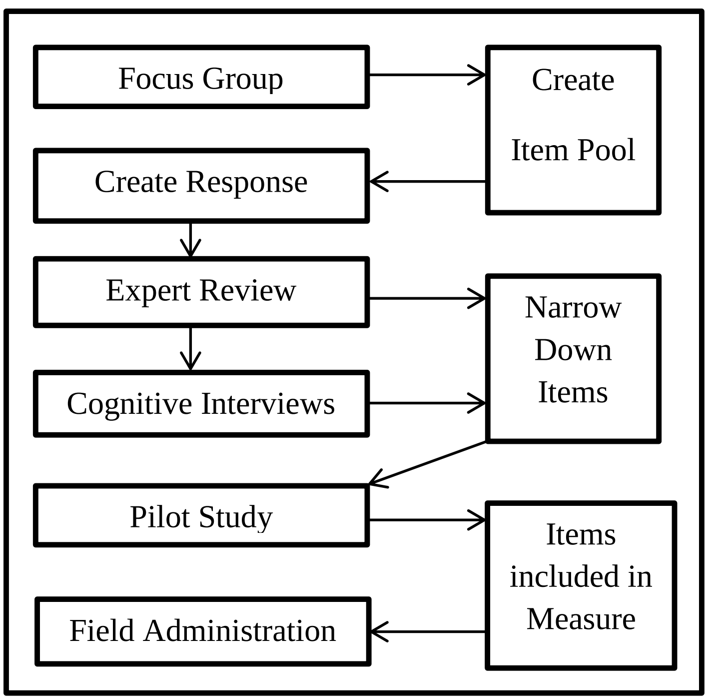


Figure 2. *Study Design Diagram*

Stage 1: Planning

In the planning stage, I determined constructs to be considered, conducted a review of the literature, identified the target population, and conducted focus groups (DeVellis, 2003). The constructs and literature review are discussed in Chapter 1. The target audience for this measure was K–12 teachers in the central United States, with the potential to be appropriate for pre-service teachers and teachers nationwide.

Focus Group.

Participants. The purpose of the focus group was to verify findings from the literature review and help guide the survey content. There were two focus groups with a total of seven teachers who represent K–12 educators. These teachers were purposively sampled to gain a variety of experiences. Participants varied by age, gender, teaching experience, and subject taught, as well as grade level experience (Table 2). I had existing relationships with these teachers, which helped set an atmosphere of trust, equality, and openness.

Table 2

Backgrounds of focus group participants

Person	Subject Expertise	Levels Taught (past/present)	Years Teaching	Sex	Relationship to Researcher
1	Language Arts and Social Studies	Elementary School	4	Female	Roommate of a friend, friendly acquaintance

2	Science and Math	Elementary School	5	Female	Taught with at the Denver Museum of Nature and Science
3	Science and Social Studies	Elementary School and Middle School	7	Female	Taught same subject and grade level at same school for 3 years
4	Language Arts and Social Studies	Middle School	8	Male	Was on the same grade level team for 2 years
5	Math and Language Arts	Middle School and High School	10	Female	Was on same grade level team for 1 year, worked closely together for 2 years in leadership teams
6	Social Studies	Middle School and High School	5	Male	Was on the same grade level team for 1 year
7	Language Arts and Social Studies	Middle School	8	Female	Wife of a friend, friendly acquaintance

Instrument. Appendix A provides the questions that guided the discussion.

Procedure. The focus groups were approximately 1.5 hours long, and food was provided. The focus groups were held in a central location to allow for a comfortable, quiet, relaxed setting. The focus groups sessions occurred on an evening during the week and a weekend afternoon, depending on the participants' convenience. Notes were taken during the sessions and these were recorded with audio only. Recording and notes were analyzed for verification of participants' thoughts, new ideas to include, and validation of existing survey ideas and the conceptual framework.

The session began by asking the participants to brainstorm answers to the following questions: "What should a teacher know about data and assessments?" and "What should a teacher be able to do with data and assessments?" Brainstorming began individually on post-its and then as a group to discuss thoughts together. These were then compared to the literature findings and discussion revolved around confirming, discounting, and adding to the framework of Knowledge→Use.

The conceptual framework from the literature review on Knowledge and Use was then explained and discussed. Specific information requested included fit, direction of relationships, and how participants could relate it to their personal experiences. Finally, I shared some of my thoughts on preliminary survey ideas asking for specific opinions and suggestions regarding survey order, survey scales, and potential items.

All of the information gathered from the focus group was analyzed to confirm, deny, and/or expand on the conceptual framework. This data guided the creation of the item pool and survey structure.

Stage 2: Construction

The construction stage was used to create, validate, and select items to be included in the final survey (DeVellis, 2003). This was accomplished by creating an item pool, deciding on appropriate response scales, asking experts to review items, conducting cognitive interviews, and reducing the number of items based on interviews.

Item Pool. The item pool included 3–4 times as many items as needed in order to pick the best options for the final survey (DeVellis, 2003). These items were narrowed down through expert review, cognitive interviews, and final construct considerations to create a more concise survey.

This process actually occurred twice. After a valuable expert interview, the survey was rethought. The initial survey concept involved two sections: Knowledge and Use. After discussion with an expert, I realized I was mixing the assessment process between the two factors. In actual teaching practice, the process is the most valuable aspect of Knowledge and Use so the measure was restructured to address this.

Response Scale. Each construct had its own response scale. The Knowledge of the Assessment Process employed a rating scale of: 1-Strongly Disagree, 2-Disagree, 3-Agree, 4-Strongly Agree. The Use of the Assessment Process was a frequency of use scale: 0-Never, 1-Yearly, 2-Quarterly, 3-Monthly, 4-Weekly, 5-Daily.

The third section of the survey was Practical Application of Different Types of Assessments. This section was not a scale. Instead teachers checked how they use different assessment types. These responses were coded 0-1.

Expert Review.

Participants. The purpose of expert review is to clarify and validate the content, structure, and items of the survey with content experts in the field of education. The six chosen experts have knowledge of how data and assessments are used in education, knowledge of psychometrics and survey design, and experience in educational research regarding assessments and/or in K–12 teaching. Six interviews were conducted with three university professors who study how teachers are using data and assessments and three PhD candidates who were former teachers and are experts in curriculum and instruction (Table 3).

Table 3

Backgrounds of Experts

Person	Background	Expertise	Sex
1	University professor	Educational research, focus on teacher effectiveness and how teachers use assessments	Male
2	University professor, professional development consultant/teacher, Colorado Department of Education	Educational research, professional development with teachers using data and assessments at the state and district level	Female
3	University professor, former math teacher	Worked with pre-service teachers, research with elementary math and science teachers on formative assessment use, former teacher insights specifically to math	Female
4	Language arts professional development coordinator, PhD candidate, former language arts teacher	Worked with teachers on using their assessments and data, former teacher insights specifically to language arts	Female
5	PhD candidate, former science and math teacher	Worked on grant studying teacher effectiveness, former teacher insights specifically to science and math	Female
6	PhD candidate, former language arts teacher	Worked on grant studying teacher effectiveness, former teacher insights specifically to teachers with “anti-data” attitudes	Male

Instrument. The interview questions and discussion are outlined in Appendix A. The response form outline can be seen in Appendix D. The response form content was modified based on the ideas seen in the focus group. There were 41 items to choose from regarding knowledge and 41 for use.

Procedure. Expert reviews were conducted in two parts: interviews and response forms. The interviews took 10–30 minutes to explain and discuss the conceptual frameworks and survey constructs (Appendix A). These were conducted in faculty offices, a coffee shop, and over the phone. After the interviews, participants were given the response form (Appendix D) and the item pool. They were asked to complete the response form immediately or at a later time. This took about 30 minutes to complete. The participants were asked to rate potential items for the survey based on representativeness of the item in the domain, clarity of the item, and item difficulty. They were also asked to give opinions on the scale appropriateness, comprehensiveness, and construct definition. Data analysis focused on suggestions from the experts, discrepancies across cases, and validation of items and of the conceptual framework. This led to modification and selection of 21 knowledge and 22 use survey items for the cognitive interviews. Items were also rephrased and organized into the assessment process framework. Multiple items were still included by topic in order to further vet them in the cognitive interview (i.e., analysis of data had seven questions at this stage and was reduced to four items in the final measure).

Cognitive Interviews.

Participants. The purpose of cognitive interviews is to verify and modify the final survey items with teachers who are within the target population but not in the final administration sample (i.e., they are in a different school district). Five cognitive interviews were conducted with a broad range of teachers in K–12 education varied by age, gender, teaching experience, and subject taught, as well as grade level taught (Table 4).

Table 4

Backgrounds of cognitive interview participants

Person	Subject Expertise	Levels Taught (past/present)	Years Teaching	Sex	Relationship to Researcher
1	Math	Middle school	10	Female	Taught same subject and grade level for 1 year
2	Math	High School	14	Male	Acquaintance from University of Denver
3	Art	Elementary School, Middle School, and High School	10	Female	Acquaintance from cycling community
4	Special Education	Middle School and High School	4	Female	Taught in same building for 1 year
5	Special Education	Elementary School and Middle School	19	Female	Taught in same building for 4 years

Instrument. Interview questions can be seen in Appendix A and survey items are seen in Appendix B. These were slightly modified based on the information gathered in the Focus Group and Expert Review and included 21 knowledge items and 22 use items.

Procedure. Interviews occurred at a convenient time and place for the participant and lasted from 60 to 90 minutes. During these interviews, I went through the survey items and discussed the interviewee's thoughts on each one. Specific information sought was a) how did he or she interpret the item, b) how was the clarity of wording, c) how would he or she answer it on the scale, and d) whether or not the scale was clear and appropriate (Appendix A). Analysis aggregated cases for agreement and discrepancies and to verify that the interpretations of items match the intended item content. This led to the final modifications of survey items for the pilot study. Items that were confusing or were interpreted in multiple manners were deleted or revised. For example, "I give students quick feedback" was debated in multiple interviews and therefore removed in favor of "I give students feedback on how to improve." Items that were repetitive were also removed with the "best" question kept based on participant feedback. For example, "I analyze information from assessments for student learning" was removed since other questions asked more specifically how this was done. Finally, teachers and experts had concerns with the difference between choosing an already existing assessment versus creating their own assessment to measure a learning objective so these two questions were combined. In the end, 15 knowledge items and 15 use items were kept.

Stage 3: Quantitative Evaluation

The quantitative evaluation stage occurred in two parts: the pilot study and the final field administration. This stage addressed the first research question:

1. What is the measured construct?
    a. Do items factor appropriately into three distinct domains of Knowledge and Use? Is the factor structure confirmed in an independent sample?
    b. Is there adequate reliability for each of these factors?
    c. Are the response scales used appropriately?
    d. Is the measure well targeted?
    e. Which items are hardest and which are easiest for the teachers to agree with?

Analysis began with item analysis in the pilot study to explore how items fit into the intended constructs and reliability analysis. The field administration then used a split sample to perform exploratory factor analysis (EFA) and Rasch analysis on one sub-sample and then confirmatory factor analysis (CFA) and Rasch analysis with the other sub-sample. Parallel analysis and EFA results were used to identify the number of factors to retain and how the items fit into factors generated using principle components analysis with varimax rotation. Only items that loaded at 0.4 or above on the appropriate factor without cross-loading (loadings differ by at least 0.1) on another factor were included (Comrey & Lee, 1992). Rasch analysis was conducted on each factor separately to again assess dimensionality and to explore scale use and how items and persons fall together. Category steps should follow the proposed order of least to greatest with even probability curves (Linacre, 2007). The item-person map was used to assess whether the items were of appropriate difficulty for the target population. This map also revealed which items were the hardest-to-easiest for teachers to agree with.

Finally, CFA was used to assess whether the proposed two-factor model for the entire measure fit the data from an independent sample. Analysis was conducted with AMOS 7.0 (Arbuckle, 2006) using raw scores. Full-information maximum likelihood estimation was used to handle missing data. Model fit was assessed using Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA) indices. CFI compares the theoretical model to a null model and is considered sufficient with values of 0.9 and above. RMSEA is sensitive to parsimony of the model with values of 0.06 and below considered acceptable.

Pilot Study.

Participants. The purpose of the pilot study was to address the first research question and modify the survey as needed before the larger field administration. The survey was piloted at a middle school in the Denver metropolitan area. This school had approximately 60 teachers with diverse backgrounds. Census sampling was used with a response rate goal of 80%. Each grade level had 10–12 teachers and each content area had around 10 teachers. While middle school grades are only part of the target population, many of these teachers had taught in elementary or high schools in the past, so this school can be considered fairly representative of the target population.

Instrument. The survey for the pilot study had two parts. Part 1 was the Knowledge and Use measure. Part 2 was demographic information. See Appendix B for the survey.

Procedure. The pilot school was chosen for convenience; I previously taught at this school and had connections with the administration and professional developer. The pilot study was done through census sampling with every teacher encouraged to participate in the survey. The survey was administered through an online link using Qualtrix and a study description sent to the teachers' school email accounts from the principal. The survey was anonymous and began with the participants reading and agreeing with the informed consent (Appendix C). The school district's Internal Review Board (IRB) was contacted and granted approval before administration of the survey.

Field Administration.

Participants. The field administration enlisted three samples. The survey was administered through an email list-serve of teachers across the state of North Dakota in May 2015 and again in August 2015. An online snowball sample was also administered during the summer of 2015 open to teachers across the United States.

Instrument. The final survey was anonymous with access through a Qualtrix link that began with informed consent (Appendix C). This survey included three parts: part 1 was the final items for the teachers' Knowledge and Use of Data and Assessment (tKUDA) measure, part 2 was the demographic information, and part 3 included 10 items from Using Data to Inform Decisions (Cavalluzzo et al., 2014) to assess convergent validity items.

Procedure. Data were collected from K–12 classroom teachers of all content areas. These teachers were selected based on a convenience sample and willingness of the North Dakota Educational Technology Council to participate. The email sent to North Dakota teachers came from the organization contact in with a description of the study and survey link. The survey was sent in May 2015 and again in August 2015. The survey was open for four weeks. The organization received a final copy of the overall analysis with areas of strengths and needs, as well as recommendations of professional development areas.

For the snowball sample, a convenience sample of K–12 teachers was found via online platforms of social media and email lists. The survey also began with a description of the study and survey link as well as encouragement to forward to other teachers they knew. This survey was open from June through September of 2015.

Stage 4: Validation

Content Validity. Content validity refers to evidence that the survey was indeed measuring the two intended, latent factors (DeVellis, 2003). This was accomplished through the opinions and modifications from the expert interviews to verify that the measure was actually determining teachers' Knowledge and Use of data and assessments. Further evidence was provided by the item-person maps generated via Rasch analysis for each of the factors. This map places items from easiest to hardest to agree with. This order was correlated to the expert review and cognitive interview ratings.

Convergent Validity. Convergent validity refers to how the measure correlates with a known measure of the same latent factor or diverges from a similar, but distinctly different latent factor (DeVellis, 2003). This was explored by adding 10 items from the Using Data to Inform Decisions survey (Cavalluzzo et al., 2014) to the field administration as Part 3. The reported reliability of this measure was $\alpha = 0.87$. Since this inventory should have measured the same construct as the Use factor, a correlation of the average Use score and the average of the Using Data to Inform Decisions scale was computed. These two constructs should be strongly correlated, which would provide support for convergent validity. Additionally, there were no appropriate scales regarding data and assessment that were relevant but distinct to assess discriminant validity.

Credibility. In order to further explore convergent validity, a qualitative method of demonstrating validity was sought (Creswell, 2007). In qualitative methods, credibility refers to establishing a connection between the constructed realities of respondents and how those realities are seen by the researcher (Sinkovics, Penz, & Ghauri, 2008). Studies are considered credible if researchers can show that their understanding and portrayal of the situation matches the participants' actual understandings. The pilot study was conducted in a school that is very familiar to me; I have firsthand knowledge of data and assessment trainings conducted there in the last seven years, and I have a close relationship with the principal and the professional developer. The results of data analysis from the pilot study were discussed with the principal and professional developer to verify and validate that their teachers responded in expected ways based on their observations and understandings of these teachers' knowledge and use of data. If the survey results from the respondents match the qualitative understandings of the principal and professional developer who know and work with these teachers daily, this gives support for credibility.

Scale Validity. Construct validity refers to how the response scales are being used (DeVellis, 2003). This was demonstrated through Rasch analysis to examine ratios between categories, test scale use, and explore category structure and function. This analysis was conducted separately for each factor since the response scales were different.

Internal Structure Validity. The validity of internal structure refers to how items group into factors and to confirm that items factor in the same manner across groups. This in part addresses research question 2, "Do teachers respond differently based on demographic differences such as gender, race/ethnicity, content expertise, and number of years teaching?" Analysis included exploratory factor analysis to determine the factor structure of the measure. Differential item functioning tested if items were answered differently across demographic groups formed by sex and racial identity.

Relationships between Knowledge and Use

After stage 4, further analysis was conducted to answer the final research question, "Is there a relationship between Knowledge and Use of data?" Correlation analysis was used to answer research question 3 using an average score for Knowledge and an average score for Use. All assumptions of correlation were tested before analysis took place using SPSS software. The significance level was set at 0.05 to test the stated hypotheses.

Research question 3 had a sub-question, “Does this relationship differ by demographic differences such as years teaching and content expertise?” that was explored by comparing independent correlations after transforming the r to Fisher’s Z. Years teaching was grouped into the following categories to compare relationships: 1–5 years, 6–15 years, 16–25 years, 26+ years. Content expertise was compared between math, science, language arts, social studies, and elective teachers. Since there are multiple independent correlations, a chi-square test for independent correlations was used to explore across groupings. According to Glass and Hopkins (1984), this equation is:

$\chi^2 = \sum w_j Z_j^2 - w.\bar{Z}_w^2,$ where $\bar{Z}_w = \frac{\sum w_j z_j}{w.}$ and $w. = \sum w_j$.

Practical Applications of Data and Assessments

In order to answer the fourth research question, “What are the practical applications of data and assessments that teachers rely on and what data sources are they using?” descriptive statistics were utilized. Descriptive statistics were used to describe what teachers reported using data for, i.e., informing instruction or giving parents feedback. Finally, I also described what sources of data teachers used, i.e., personal formative data, personal summative data, district data, or state data. Spearman rank order correlations were used to explore relationships between frequency of reported use by assessment type and purpose of that use.

CHAPTER 3: RESULTS

This chapter describes the results of the pilot study and the field administration of the Teachers' Knowledge and Use of Data and Assessment (tKUDA) measure for K–12 teachers. The pilot study was conducted to understand how the measure works within a single school setting in order to refine items, verify scale appropriateness, and explore credibility of results. The field administration was conducted to explore results of teachers' tKUDA responses, including reliability and validity of the measure, and to understand dimensionality, scale use, and item function. The field administration data were collected via three surveys in order to create a split sample for exploratory and confirmatory analysis of measure structure. Surveys were merged for analysis of variance, regression, and chi-square analyses.

Pilot Administration

Procedure. A middle school was chosen from the Denver-metropolitan area. This convenience sample was chosen because the teachers and administrators know me and were open to the study. District IRB approval was received prior to the survey administration. Teachers were sent an email with the online survey link. The survey was open for three weeks with reminder emails sent from the professional development personnel once a week. They could complete the survey at any convenient time within the three weeks. During the first week, I was able to walk around during teachers' planning time and professional learning team meetings to explain the survey's purpose, answer questions/concerns, and encourage participation.

Instrument. The first question involved consent to participate in the survey (Appendix C). If "I agree to participate" was chosen the Qualtrics online survey moved them to the first question, while if "I do not agree to participate" was chosen they were exited from the survey. Part one of the survey included 15 Knowledge items and part two contained 15 Use items. Part three asked teachers what assessment types they used for 22 practical applications of data. Part four comprised demographic items on years teaching, race/ethnicity, sex, content expertise, and grade level taught. See Appendix B for more detail.

Participants. Thirty-four teachers responded with full survey information. Six other partial surveys were discarded. Of the 56 teachers available to take the survey, the response rate for complete surveys was 61%. Demographic information of the respondents can be seen in Table 5.

Table 5
Pilot Administration Demographics

Race/Ethnicity	Gender	Grade Level	Content Expertise
89.7% White 6.9% African-American 3.4% Hispanic	17.2% Male 72.4% Female 3.4 % Other 6.9% Prefer not to answer	22.2% 6^{th} 25.9% 7^{th} 22.2% 8^{th} 29.6% all	24.1% Math 31.0% Science 10.3% Language Arts 13.8% Social Studies 3.4% Art 3.4% Other 13.8% Special Education

Analysis. Analysis began by examining responses to each survey part separately, starting with the demographic information. New variables were created for the Knowledge and Use factors that averaged scores from the 15 items in each scale. Partial responses were removed, as teachers who needed to stop and restart the survey reported being unable to restart in the same place; they instead started over. While it is important to note that the following analysis needs to be considered tentative with only 34 responses, this was valuable information for the school itself with 61% of their teachers giving voice to their understanding of data and assessment.

Knowledge of the Assessment Process. The Knowledge scale employed a rating scale with 1 = Strongly Disagree and 4 = Strongly Agree. With a mean score across the 15 items of 3.12, teachers were fairly confident in their knowledge of the assessment process, regardless of whether or not they were able to actually do these tasks in their practice. In comparing means across items, these teachers were very knowledgeable in regard to setting goals (3.50), communicating goals to students (3.50), assessing prior knowledge (3.40), and using a variety of assessment techniques (3.40). They appeared less comfortable in their knowledge of giving students feedback (3.15), reflecting on their practice via assessments (3.15), and revising instruction based on assessments (3.18). Additionally they responded with lower agreement on reading disaggregating data (3.15) and disaggregating their own data (3.09). Based on my experiences in the school and from the literature, I expected a higher frequency of disagreement on these items. Only 6 of the 34 teachers responded with Disagree or Strongly Disagree on the items about data use.

Reliability and item analysis were conducted with the Knowledge factor. Cronbach's alpha was very high at 0.96. All items fit a Knowledge factor very well with item-total correlations between 0.67 and 0.89. Results can be seen in Table 6.

Table 6
Pilot Administration Knowledge Item Statistics

Item	Mean	Std. Dev.	N	Corrected Item-Total Correlation
1. I know how to set specific learning goals/objectives.	3.45	.72	31	.78
2. I know how to communicate my learning goals to students using multiple methods.	3.55	.72	31	.67
3. I know how to assess my students for prior knowledge.	3.42	.67	31	.89
4. I know how to choose or create an assessment strategy that will measure my specific learning objective.	3.26	.73	31	.70
5. I know how to use a variety of assessment techniques.	3.39	.72	31	.79
6. I know how to effectively use assessments to show students' thinking, not just their answers.	3.03	.75	31	.77
7. I know how to evaluate evidence from assessments in order to prove student learning.	3.03	.79	31	.79
8. I know how to evaluate evidence from assessments in order to improve student learning.	3.10	.70	31	.81
9. I know how to read data (typically from a standardized exam) when it is broken down for me based on gender, race/ethnicity, IEP, GT, content standard, etc.	3.19	.75	31	.71
10. I know how to break down results from my own assessments based on gender, race/ethnicity, IEP, GT, content standard, etc.	3.06	.77	31	.64
11. I know how to give students specific feedback on what they need to improve.	3.23	.67	31	.85

12. I know how to give students specific feedback on how to improve.	3.13	.72	31	.79
13. I know how to reflect on my instructional practices based on evidence from my assessment techniques.	3.23	.62	31	.75
14. I know how to revise my instructional practices immediately (on the fly) based on evidence from assessments.	3.19	.70	31	.76
15. I know how to revise my instructional practices for the next year based on evidence from assessments.	3.19	.60	31	.79

Use of the Assessment Process. The Use items had a response scale of 1-Never, 2-Yearly, 3-Quarterly, 4-Monthly, 5-Weekly, and 6-Daily; higher values indicate more frequent use. With an overall mean of 4.44, teachers reported using the assessment process on a consistent monthly basis. Once again, setting goals (5.48) and communicating goals (5.26) were the most frequent uses of assessment. Assessing prior knowledge (4.63), matching assessments to goals/standards (4.38), using a variety of assessments (4.68), and showing students' thinking (4.38) all seemed to happen monthly. Reading disaggregated data (3.19) and disaggregating their own data (3.19) appeared to occur quarterly. Giving students' feedback (4.66) happens on a monthly to weekly basis, but Knowledge values for this item was low. Reflecting (5.03) and revising instruction immediately (5.23) were frequently done but also had lower values for Knowledge. It appears that teachers are giving feedback, reflecting, and revising instruction quite frequently, but are not as confident in their knowledge of how to go about it. Finally, teachers reported that they are revising instruction for the next year (3.67) with the lowest frequency.

Item analysis and reliability were also estimated for the Use items. The estimated Cronbach's alpha for Use was 0.91 which is lower than the Knowledge factor but still high. Most items fit this factor well with item-total correlations between 0.49 and 0.79. Three items had lower correlations than desired, and these may be problematic: I typically communicate my learning goals to students using multiple methods ($r = 0.43$); I break down results from my own assessments based on gender, race/ethnicity, IEP, GT, content standard, etc. ($r = 0.42$); and I revise my instructional practices for the next year based evidence from assessment techniques ($r = 0.34$). Since deleting these items only raised the reliability to 0.92, items were kept for field administration with special attention to these items given in the next analysis. Results are seen in Table 7.

Table 7
Pilot Administration Use Item Statistics

Item	Mean	Std. Deviation	N	Corrected Item-Total Correlation
1. I set specific learning goals/objectives…	5.41	1.12	27	.62
2. I typically communicate my learning goals to students using multiple methods…	5.15	1.38	27	.43
3. I assess my students for prior knowledge	4.74	1.29	27	.49
4. I choose or create an assessment strategy to measure a specific learning goal…	4.41	1.22	27	.75
5. I use a variety of assessment techniques…	4.59	1.25	27	.73
6. I effectively use assessments to show students' thinking, not just their answers…	4.30	1.35	27	.68
7. I evaluate evidence from assessments in order to prove student learning…	4.22	1.16	27	.73
8. I evaluate evidence from assessments in order to improve student learning…	4.26	1.23	27	.78

9. I typically read data that is broken down for me based on gender, race/ethnicity, IEP, GT, content standard, etc. ...	3.11	1.05	27	.54
10. I break down results from my own assessments based on gender, race/ethnicity, IEP, GT, content standard, etc. …	3.04	1.09	27	.42
11. I give students specific feedback on what they need to improve…	4.63	1.15	27	.61
12. I give students specific feedback on how to improve…	4.67	1.14	27	.63
13. I reflect on my instructional practices based on evidence from assessment techniques…	5.07	1.17	27	.71
14. I revise my instructional practices immediately (on the fly) based on evidence from assessment techniques…	5.22	1.09	27	.79
15. I revise my instructional practices for the next year based on evidence from assessment techniques…	3.59	1.19	27	.34

Application of Different Assessment Types. These items asked teachers to check all types of data that apply for that application, such as informing instruction. A great deal of specific information was generated by this section of the measure, but only highlights are discussed here.

Teachers rely heavily on their own personal formative and summative assessments and much less on district and state assessments. Overall there appears to be less emphasis on creating student groups, gauging student engagement, get a "feeling for" incoming students, student goal setting, disaggregating data, and predicting scores using a variety of data. Data still seem to be used for these applications, but in a more specific manner. For example, gauging student engagement derives more from formative assessments while disaggregating data happens using summative assessments. On the other hand, there appears to be more emphasis on informing instruction, giving student and parent feedback, and identifying student strength/weakness using a variety of data.

Credibility. While validity analysis is not possible with only 34 teachers, I wanted to understand if there was any credibility I could establish from the tKUDA measure. Since all the respondents are from the same school I interviewed the professional developer, who will be called Carina for this report. She has been at the school for over 10 years and has been in multiple leadership positions, including serving as the school's professional developer for the last two years. Before the survey was deployed we discussed what she thought the results would be and then met afterward to discuss findings. I was also able to give her recommendations for the next school year on potential school-wide trainings. The following provides an additional insight into the quantitative data. Discussions with Carina help demonstrate that the survey responses seemed to be credible; the numbers seemed to be honestly reported and responses matched what the professional developer had seen in practice.

Knowledge of the Assessment Process. The fact that the setting and communicating goals items had high means was no surprise to Carina. There was a school-wide expectation for teachers to write, post, and review the target learning goals for that day or week. My surprise at the level of knowledge around disaggregating data was also discussed. Carina was very happy with this agreement across teachers, as it has been a major focus in her teacher trainings this year. New data trainings and resources became available and she has been working school-wide and with specific groups to become more adept at reading and interpreting disaggregated data and finding ways that teachers can actually break down their own data into categories.

Use of the Assessment Process. There were no surprises from Carina and she was able to provide possible explanations for the item means seen. The school-wide expectation for goal setting led to the weekly to daily frequency seen. The items regarding choosing assessments to match student thinking and goals/standards were a little lower than expected. Carina thought this was due to the new common core standards, the new state exam, and many teachers teaching new content. There was uncertainty around content standards and how they would be assessed during the 2014–2015 school year as Colorado was transiting to the new state assessment. Items on disaggregating data showed teachers doing this quarterly which was confirming as the data trainings had focused on using quarterly benchmark tests. The discrepancy between Knowledge and Use for feedback was concerning to Carina. She had been observing the feedback that teachers were giving to students and had already decided to focus on professional development opportunities on giving students feedback for the next year. She believed that teachers were trying to give feedback to the best of their ability, but more techniques and guidance were needed. Finally, the low frequency of revising instruction for the next year could be explained by the fact that teacher turnover rate in the school is high and even those that stay are moved into new content areas and different grade levels each year. Carina thought they may not invest their time to contemplate revisions for the next year if they were unsure what they would be teaching.

Conclusion and Modifications for Field Administration. The overarching conclusion was that the tKUDA measure was generally working as intended. All items were kept for field administration with no modifications or additions. Only two changes were made for field administration.

The first change was to the response scale. Teachers reported verbally and in a written response that they struggled with only having Agree or Strongly Agree options in the Knowledge scale; they wanted a Somewhat Agree choice. Since the Disagree and Strongly Disagree options were both used, but with low frequency, I was not comfortable discarding these choices but instead chose to merge them. Not surprisingly, it appears that teachers have more agreement in their Knowledge of the assessment process. In order to increase potential variability in responses the new response scale for the Knowledge scale was: Disagree, Somewhat Agree, Agree, and Strongly Agree.

The second change was to the online platform for the survey. Teachers were trying to take the survey during their "free" time, meaning before or after school, at lunch, or during their planning periods. Teachers are always interrupted during these times by colleagues, students, or meetings. The survey needed to have the option to start and stop, while retaining the previous information so that teachers could come back to it at their next convenience. This was ensured in the online Qualtrics survey before field administration was given.

Field Administration

Once these changes were incorporated, the survey was ready for data collection via field administration in North Dakota and the snowball sample across the United States.

Analysis of Measure Structure

Principal components analysis (PCA) was conducted on the Knowledge and Use factors separately. When all items were combined in the pilot study and with this sample, items factored into distinctly separate factors (Appendix E). New variables were created for the Knowledge and Use factors that averaged scores from the 15 items in each scale. Pairwise deletion was used to accommodate missing data. PCA was used with the first sample, called the calibration sample. Confirmatory factor analysis (CFA) was conducted with the second sample, called the validation sample. Rasch analysis was conducted with both the first and second samples to examine measure structure.

Research questions addressed by this analysis were: What is the measured construct? Do items factor appropriately into distinct domains of Knowledge and Use? Is there adequate reliability and validity for each of these factors?

There were 201 teachers who responded to the calibration survey. A description of demographic information for the respondents in the first sample can be seen in Table 8 below.

Table 8
Calibration Sample Participant Demographics

Race/Ethnicity	Gender	Grade Level	Content Expertise
90.8% White	19.3% Male	14.3% K–2nd	20.7% Math
1.4% Am. Indian/Alaska Native	80.7% Female	18.6% 3rd–5th	7.9% Science
0.7% Hispanic/Latino		27.1% 6th–8th	20.7% Lang. Arts
2.1% Other		17.9% 9th–10th	13.6% Social Studies
3.5% Prefer not to answer		22.1% 11th–12th	5.0% Music
			0.7% Art
			0.7% Physical Ed
			3.6% Computer/Tech
			1.4% Foreign Lang
			13.8% Special Ed
			12.1% Other

Knowledge of the Assessment Process. The Knowledge items had a rating scale with 1- Disagree, 2-Agree, 3-Somewhat Agree, and 4-Strongly Agree. With a mean score across the 15 items of 3.20, teachers were confident in their knowledge of the assessment process, regardless of if they were able to actually do these tasks in their practice. Comparing item means, these teachers felt very knowledgeable in regard to setting learning goals (3.43) and communicating these objectives to students (3.38). They were less comfortable in their knowledge of reading disaggregated data (3.02) and disaggregating their own data (2.85). The mode across all Knowledge items was 3.0.

The category of Disagree was not used for the following items: setting learning objectives, assessing prior knowledge, choosing assessments based on objectives, using a variety of assessment techniques, giving feedback on what and how to improve, reflecting using data, and revising now and later. Results can be seen in Table 9. Reliability estimation and item analysis were conducted with the Knowledge factor. Cronbach's alpha was high at 0.94.

Table 9

Knowledge of the Assessment Process Item Statistics

Item	Mean	Min-Max	N	Mode	Factor Loading
Setting learning objectives	3.43	2–4	200	3	.74
Communicating objectives	3.38	1–4	200	3	.60
Assessing prior knowledge	3.28	2–4	200	3	.77
Choosing assessment from objective	3.31	2–4	200	3	.75
Using a variety of assessments	3.29	2–4	199	3	.77
Showing student thinking	3.06	1–4	198	3	.74
Using assessments to prove learning	3.04	1–4	199	3	.76
Using assessments to improve learning	3.11	1–4	198	3	.80
Reading disaggregated data	3.02	1–4	198	3	.58
Disaggregating their own data	2.85	1–4	200	3	.64
Giving feedback on what to improve	3.30	2–4	200	3	.76

Giving feedback on how to improve	3.19	2–4	197	3	.76
Reflecting using data	3.27	2–4	199	3	.79
Revising instruction immediately	3.24	2–4	199	3	.78
Revising instruction later	3.31	2–4	198	3	.82

Bartlett's test of sphericity and the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy were analyzed to assess the suitability of the data for factor analysis. The results revealed that KMO = 0.92 and Bartlett's test was significant at $p < 0.001$ with a χ^2 (105) = 1944.38, demonstrating suitability to conduct PCA. Comrey and Lee's (1992) guidelines for factor loadings were used as cut off values for this study: 0.71 is excellent, 0.63 is very good, 0.55 is good, 0.45 is fair, and anything below 0.32 is poor. All items fit the Knowledge factor with factor loadings between 0.58 and 0.82. Results can be seen in Table 9 above.

Dimensionality was assessed using eigenvalues, examination of the scree plot, and parallel analysis. One overarching factor was retained based on an eigenvalue of 8.22 and examination of the scree plot (Figure 3). A possible small second factor is seen with an eigenvalue of 1.20. Parallel analysis was run to see if the decision about retaining one factor was justified. The first eigenvalue from the parallel analysis was 1.58 which is less than the eigenvalues from the PCA of 8.22. The second eigenvalue from the parallel analysis was 1.44 which is more than the PCA value of 1.20. All other parallel analysis eigenvalues were greater than eigenvalues from the PCA run which gives support for a single factor. Using a varimax rotation, the conceptual definition of two factors were explored. Multiple items crossloaded on both factors but the strongest items on the second factor were the two on disaggregating data. This demonstrates that the small potential second factor was due to the analysis questions, but when forcing one factor all items loaded above 0.58. Based on these results, the Knowledge factor was considered unidimensional.

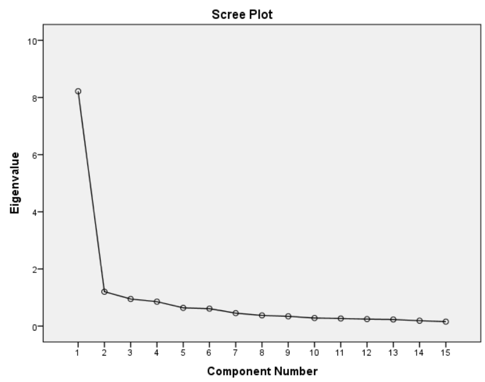


Figure 3. *Scree Plot for Knowledge of Assessment Process*

Use of the Assessment Process. The Use items had a rating scale of 1-Never, 2-Yearly, 3-Quarterly, 4-Monthly, 5-Weekly, and 6-Daily. With a mean score across the 15 items of 4.54 teachers use data monthly to weekly. Comparing item means, these teachers reported the most frequent use of data when giving feedback (5.13, 5.16), reflecting (5.04), and revising instruction immediately (5.22). The least frequent use of data was regarding disaggregation of data (3.00, 2.88). Interestingly, the mode for disaggregating their own data was 1-Never. Results can be seen in Table 10. Reliability estimation and item analysis was conducted with the Use factor. Cronbach’s alpha was high at 0.86.

Table 10

Use of the Assessment Process Item Statistics

Item	Mean	Min-Max	N	Mode	Factor Loading
Setting learning objectives	4.91	1–6	173	5	.66
Communicating objectives	4.83	1–6	173	6	.62
Assessing prior knowledge	4.77	1–6	173	5	.50
Choosing assessment from objective	4.77	1–6	171	5	.64
Using a variety of assessments	4.99	1–6	172	5	.58
Showing student thinking	4.62	1–6	172	5	.64
Using assessments to prove learning	4.41	1–6	172	5	.69
Using assessments to improve learning	4.54	1–6	170	5	.70
Reading disaggregated data	3.00	1–6	172	3	.54
Disaggregating their own data	2.88	1–6	169	1	.53
Giving feedback on what to improve	5.13	1–6	172	6	.56
Giving feedback on how to improve	5.16	1–6	171	6	.64
Reflecting using data	5.04	1–6	170	6	.67
Revising instruction immediately	5.22	1–6	172	6	.47
Revising instruction later	3.73	1–6	170	2	.48

Bartlett’s test of sphericity and the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy were analyzed to assess the suitability of the items for factor analysis.

The results revealed that KMO = 0.82 and Bartlett's test was significant at $p < 0.001$ with a χ^2 (105) = 934.81, demonstrating suitability to conduct a PCA.

Most items demonstrated good fit to a single dimension with factor loadings between 0.70 and 0.47. Results can be seen in Table 10 above.

Dimensionality was assessed first. Four factors were possible based on eigenvalues above 1.0 and examination of the scree plot (Figure 4). There was one primary factor with an eigenvalue of 5.38 and three possible small factors with eigenvalues of 1.54, 1.25, and 1.16. Parallel analysis was run to see if the decision about retaining one factor was justified. The first three eigenvalues from the parallel analysis were 1.58, 1.44, and 1.35, respectively, while the eigenvalues from the PCA were 5.38, 1.25, and 1.16. Only one factor was indicated by the results of the parallel analysis. All other parallel analysis eigenvalues were greater than eigenvalues from the PCA. Looking at factor loadings in the unrotated matrix, all items loaded between 0.47 and 0.70 giving evidence of one factor. While there do seem to be smaller facets of the whole Use factor, this measure can be considered fairly unidimensional.

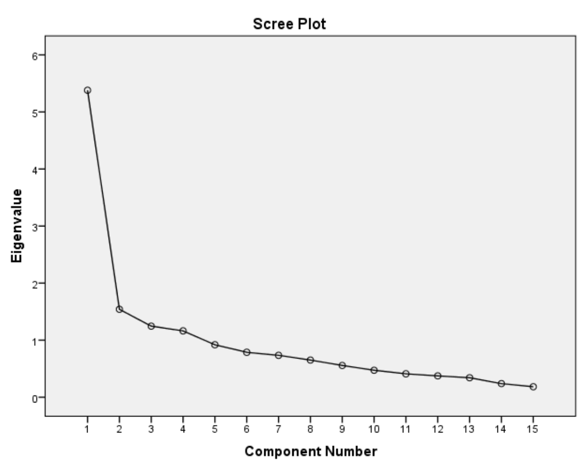


Figure 4. *Scree Plot for Use of Assessment Process*

Rasch Analysis

Rasch (1980) analysis was conducted to further examine the structure of the tKUDA scale using the software, Winsteps (Linacre 2015). A major assumption of the Rasch model is that the scale measures a unidimensional construct. Therefore, two separate Rasch analyses were conducted, with Knowledge of the Assessment Process and Use of the Assessment Process items analyzed separately. These analyses were conducted using two samples, one to calibrate the measures and the other to validate those findings with an independent sample. There were 201 teachers who responded to the calibration survey who are described in Table 8 (above with EFA results) and 164 respondents to the validation survey described in Table 17 (below with CFA results).

Research questions addressed by these analyses were: Are item sets unidimensional? Is there adequate reliability and validity for each of these factors? Are the response scales used appropriately? Is the measure well targeted? Which items are the hardest and which are easiest for the teachers to agree with?

Knowledge of the Assessment Process

Dimensionality. Linacre's (2015) suggestion for evaluation of unidimensionality is to use a principle components analysis of residuals. If the measure explains 40% or more of the total raw variance, with the first contrast (equivalent to a second factor) having an eigenvalue of 2.0 or less with less than 5% variance due to the first contrast, Linacre considers the item set to be relatively unidimensional. For the calibration sample, the measure explained 48.1% of the variance with the unexplained variance in the first contrast having an eigenvalue of 2.17 with 8.4% unexplained variance. In the validation sample, the measure explained 45.8% of the variance with the unexplained variance in the first contrast having an eigenvalue of 2.04 with 7.8% unexplained variance. The eigenvalue is slightly higher than the expectation, but this is quite common for short measures. Therefore, this factor met the expectations of unidimensionality.

Overall fit was examined as further evidence of unidimensionality. Specifically, mean square (MNSQ) infit and outfit statistics were examined. Linacre (2015) says the mean MNSQ infit and outfit should be close to 1.0. The data fit the model in the calibration sample with an average infit MNSQ of 0.94 (*SD*= 0.77) and an average outfit MNSQ of 0.98 (*SD*= 0.99) as well as in the validation sample with infit MNSQ of 0.93 (*SD*= 0.68) and an average outfit MNSQ of 0.92 (*SD*= 0.74). These values indicate that there was some overall unexpected behavior seen, with the data fitting the model a little too well.

As seen in Table 11, there was no unexpected behavior with respect to dimensionality or model fit. Table indices are described in some detail below.

Table 11
Knowledge Factor Dimensionality, Fit, and Separation

Index	Calibration Sample	Validation Sample
Dimensionality – eigenvalue for 1st contrast	2.17	2.04
Mean MNSQ Infit	.94	.93
SD MNSQ Infit	.77	.68
Mean MNSQ Outfit	.98	.92

SD MNSQ Outfit	.99	.74
Real Person Separation	2.52	2.79
Real Person Root Mean Square Error	.90	.81
Real Reliability of Person Separation	.86	.89
Cronbach's Alpha	.95	.95
Real Item Separation	4.14	3.05
Real Reliability of Item Separation	.94	.90

Note. Mean MNSQ Infit measures average deviation from the model providing sensitivity to midrange observations. Mean MNSQ Outfit measures deviation from the model providing sensitivity to extreme responses. Real Person/Item Separation is the ratio of the true standard deviation to the error standard deviation. Real Person Root Mean Square Error is standard error of the measure inflated for misfit. Real Reliability of Person/Item Separation = $Separation^2 / (1 + Separation^2)$.

Item and Person Fit. Item fit was examined to ensure that each item fit the Rasch model. According to Wright and Linacre (1994), the infit MNSQ values for a rating scale should be between 0.6 and 1.4 (Table 12). The MNSQ infit for items in the samples ranged from 0.80 to 1.23 (Table 12). Items 2, 9, and 10 had infit MNSQ values above 1.40 and were therefore deleted from both samples. Dimensionality results reported above reflect the final model without these items. Item separation for the calibration sample was 4.14 and 3.05 for validation sample; these values should be above 2.0 with higher values being more desirable (Linacre, 2015). Reliability of item separation was 0.94 and 0.90 for the two samples (Table 11).

Table 12
Knowledge Factor Item Fit Statistics

Calibration Sample				
Item#	Logit Position	SE	Infit MNSQ	Pt-Measure Correlation
1	-1.49	.20	1.06	.71
3	-.38	.20	.85	.76
4	-.53	.20	1.01	.75
5	-.40	.20	.94	.79
6	1.40	.20	1.23	.78
7	1.55	.19	.94	.79
8	-.94	.20	.94	.81
11	-.49	.20	.94	.77
12	.31	.20	1.18	.79
13	-.22	.20	1.05	.78
14	-.07	.20	1.05	.79
15	-.61	.20	.85	.82
Validation Sample				
Item#	Logit Position	SE	Infit MNSQ	Pt-Measure Correlation
1	-1.02	.21	.80	.80
3	-.57	.21	1.09	.76
4	-.53	.21	.89	.81
5	-.84	.21	.96	.79
6	.88	.20	1.11	.80
7	.66	.20	.89	.81
8	.95	.20	.94	.82
11	-.26	.21	.84	.78
12	.74	.20	1.09	.75
13	-.16	.21	1.10	.78
14	.54	.20	1.22	.82
15	-.39	.21	.99	.78

Person fit was also examined to ensure that individuals were answering in a consistent, expected manner. Linacre's (2015) recommendation for person fit is MNSQ infit under 4.0. One teacher was deleted from the calibration sample due to MNSQ infit over 4.0; that case underfit the model. This person's scores were deleted from the sample and the model was rerun. All other teachers fit the model well with no MNSQ infit values over 3.67. The validation sample had no person misfit. All tables above reflect the final model with this teacher removed.

Reliability. Reliability is measured by computing person and item spread across the measure. Person separation explores the ability of items to identity levels of the measure across persons on a less-to-more continuum (Bond & Fox, 2001). A separation of 2.0 is considered minimal with higher levels of separation indicating a wider range of items and persons (Linacre, 2015). Person separation for the calibration sample was 2.52, with reliability of person separation of 0.86, and a Cronbach's alpha of 0.95. The validation sample had slightly higher reliability with person separation of 2.82 and reliability of person separation of 0.89 (Table 11).

Scale Use. Results of the Rasch analysis indicated that teachers in this sample used the rating scale as intended (Table 13). Category probability curves (Figure 5) indicate an even distribution of the four categories with clearly advancing steps. Rasch-Andrich thresholds increased with category values with no evidence of step misfit with MNSQ infit values under 2.0 (Linacre, 2015). Teachers used the third category of "Agree" more frequently, suggesting our sample was very agreeable in their knowledge, but many could not confidently say "Strongly Agree." The validation sample's scale use was very similar with no unexpected behavior.

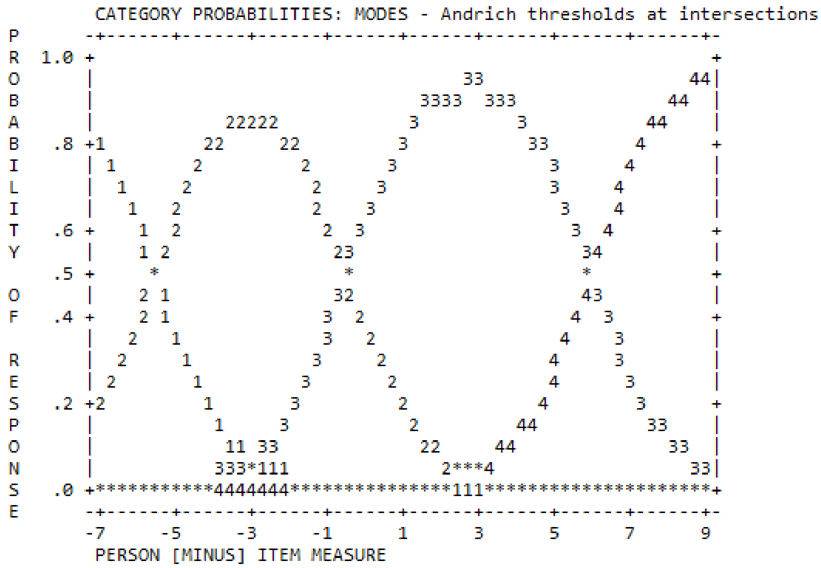


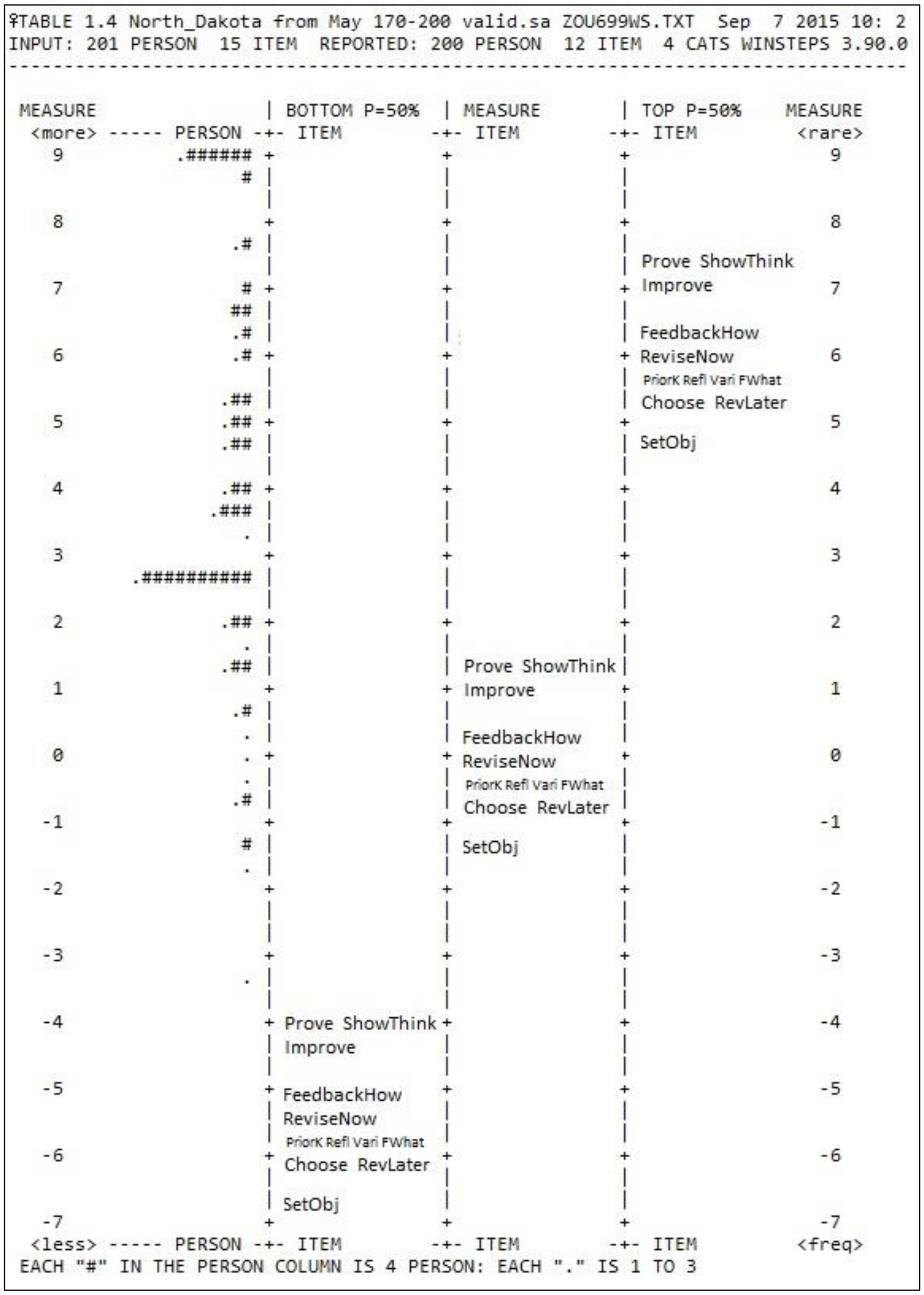
Figure 5. *Knowledge Factor Category Probability Curves*
*Categories: 1-Disagree, 2-Somewhat Agree, 3-Agree, 4-Strongly Agree

Table 13
Knowledge Factor Step Structure

Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
Calibration Sample				
1 Disagree	<1%	-2.74	1.23	(-6.61)
2 Somewhat Agree	8%	-.20	1.00	-3.00
3 Agree	59%	3.21	1.02	2.75
4 Strongly Agree	32%	6.45	.98	(7.10)
Validation Sample				
Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
1 Disagree	1	-4.00	1.16	(-5.020
2 Somewhat Agree	8	-.45	.95	-2.50
3 Agree	55	2.45	.97	2.03
4 Strongly Agree	36	5.41	1.03	(6.11)

Note. Observed percentage is the percent of all responses for that category. Observed average is the average of the measure to produce the responses observed in the category. Infit MNSQ is the average of the infit MNSQs associated with responses in that category. Step Structure is the logit position where the transition is made from a lower to this category.

Targeting and Construct Coverage. The item-person map provided in Figure 6 presents items and persons on the same scale and demonstrates scale functioning for the calibration sample. Participants represented near the top of the left-hand side of the item-person map are teachers who have higher levels of knowledge of the assessment process; participants represented near the bottom are teachers who scored lower on their reported knowledge. Participants were spread fairly evenly throughout the item-person map, although the majority of the sample appears near the top. Representation of items and participants along the map suggest this sample of teachers report strong knowledge of the assessment process. The validation sample demonstrated the same pattern with no unexpected differences in item ordering.



```
TABLE 1.4 North_Dakota from May 170-200 valid.sa ZOU699WS.TXT  Sep  7 2015 10: 2
INPUT: 201 PERSON  15 ITEM  REPORTED: 200 PERSON  12 ITEM  4 CATS WINSTEPS 3.90.0
--------------------------------------------------------------------------------

MEASURE               | BOTTOM P=50%  | MEASURE        | TOP P=50%     MEASURE
 <more> ----- PERSON -+- ITEM        -+- ITEM         -+- ITEM          <rare>
   9          .###### +               +                +                  9
                    # |               |                |
                      |               |                |
   8                  +               +                +                  8
                   .# |               |                |
                      |               |                | Prove ShowThink
   7                # +               +                + Improve          7
                   ## |               |                |
                   .# |               |                | FeedbackHow
   6               .# +               +                + ReviseNow        6
                      |               |                | PriorK Refl Vari FWhat
                  .## |               |                | Choose RevLater
   5              .## +               +                +                  5
                  .## |               |                | SetObj
                      |               |                |
   4              .## +               +                +                  4
                 .### |               |                |
                    . |               |                |
   3                  +               +                +                  3
          .########## |               |                |
                      |               |                |
   2              .## +               +                +                  2
                    . |               |                |
                  .## |               | Prove ShowThink|
   1                  +               + Improve        +                  1
                   .# |               |                |
                    . |               | FeedbackHow    |
   0                . +               + ReviseNow      +                  0
                    . |               | PriorK Refl Vari FWhat
                   .# |               | Choose RevLater|
  -1                  +               +                +                 -1
                    # |               | SetObj         |
                    . |               |                |
  -2                  +               +                +                 -2
                      |               |                |
                      |               |                |
  -3                  +               +                +                 -3
                    . |               |                |
                      |               |                |
  -4                  + Prove ShowThink +              +                 -4
                      | Improve       |                |
                      |               |                |
  -5                  + FeedbackHow   +                +                 -5
                      | ReviseNow     |                |
                      | PriorK Refl Vari FWhat         |
  -6                  + Choose RevLater +              +                 -6
                      |               |                |
                      | SetObj        |                |
  -7                  +               +                +                 -7
 <less> ----- PERSON -+- ITEM        -+- ITEM         -+- ITEM          <freq>
 EACH "#" IN THE PERSON COLUMN IS 4 PERSON: EACH "." IS 1 TO 3
```

Figure 6. *Knowledge Item-Person Map*

Rasch analysis allows researchers to see how items act on a continuum of "easiest" to agree with to "hardest" to agree with, thus creating a ruler of construct coverage. The items at the top of Figure 6 are questions that were hardest for participants to agree with, while those at the bottom were the easiest to agree with. The distribution of items, seen on the right-hand side of Figure 2, had item logit values roughly between -1.5 and 1.5, indicating the tKUDA is a good measure along this range (Linacre, 2015). Items are not spread well; the ruler is very short and could be expanded, specifically to capture higher knowledge by adding questions that are more difficult to agree with. There are three areas where the measure could be improved. Items 7 ("prove learning") and 6 ("show thinking") are equal on the "ruler" suggesting that only one of these two items is needed and the second is redundant as it is a measure of the construct at the same location. Item 4 ("choose assessments") and item 15 ("revise later") had equivalent item positions as well. Finally there were four items that fell together: item 3 ("assessing prior knowledge"), item 13 ("reflection"), item 5 ("using a variety of assessments"), and item 11 ("giving feedback on what to improve").

Invariance. Differential item functioning (DIF) was examined to ensure that the items were functioning in the same way across demographic groups. DIF is assessed using t-tests of the significance of differences in item logit position. At a significance level of 0.01, there was no DIF in the calibration sample between race/ethnicity (coded white/minority), grade level taught, or content expertise (collapsed to math, science, language arts, social studies, special education, and electives). Responses to the tKUDA were invariant across those variables. DIF was assessed for sex and was significant for item 12 ("giving feedback on how to improve") at $p = 0.002$ and for item 14 ("revising instruction immediately") at $p < 0.001$, with females scoring higher on both items. The validation sample showed no DIF for race/ethnicity, grade level, content expertise, or sex.

Summary. The Knowledge of the Assessment Process measure can be considered unidimensional based on both PCA and Rasch analyses. Items 2, 9, and 10 showed misfit and needed to be removed. Item and person separation were acceptable and items were nicely spread throughout the continuum. The measure showed support for reliability with a Cronbach's alpha of 0.95, reliability of item separation of 0.92, and a reliability of person separation of 0.86. The scale of Disagree, Somewhat Agree, Agree, and Strongly Agree was used as intended. Items spread could be improved by deleting overlapping items and expanding the scale with harder to agree with items. The measure can be considered invariant across grade level, content, and race but not necessarily sex.

Use of the Assessment Process.

Dimensionality. For the calibration sample, the measure explained 42.4% of the variance with the unexplained variance in the first contrast having an eigenvalue of 1.72 with 6.6% unexplained variance. In the validation sample, the measure explained 42.8% of the variance with the unexplained variance in the first contrast having an eigenvalue of 1.81 with 7.0% unexplained variance. This is evidence of unidimensionality.

Overall fit was examined as further evidence of a unidimensional model. Specifically, mean square (MNSQ) infit and outfit statistics were examined. The data fit the model in the calibration sample with an average infit MNSQ of 0.91 (*SD*= 0.75) and an average outfit MNSQ of 0.91 (*SD*= 0.73) and in the validation sample with an average infit MNSQ of 0.92 (*SD*= 0.89) and an average outfit MNSQ of 0.95 (*SD*= 1.08). These values indicate that while was some overall unexpected behavior seen, a unidimensional model was supported.

Table 14
Use Factor Dimensionality, Fit, and Separation

Index	Calibration Sample	Validation Sample
Dimensionality – eigenvalue for 1st contrast	1.72	1.81
Mean MNSQ Infit	.91	.92
SD MNSQ Infit	.75	.89
Mean MNSQ Outfit	.91	.95
SD MNSQ Outfit	.73	1.08
Real Person Separation	1.11	.94
Real Person Root Mean Square Error	.82	.89
Real Reliability of Person Separation	.55	.47
Cronbach's Alpha	.96	.94
Real Item Separation	4.39	4.07
Real Reliability of Item Separation	.95	.94

Note. Mean MNSQ Infit measures average deviation from the model providing sensitivity to midrange observations. Mean MNSQ Outfit measures deviation from the model providing sensitivity to extreme responses. Real Person/Item Separation is the ratio of the true standard deviation to the error standard deviation. Real Person Root Mean Square Error is standard error of the measure inflated for misfit. Real Reliability of Person/Item Separation = $\text{Separation}^2 / (1 + \text{Separation}^2)$.

Item and Person Fit. Item fit was examined to ensure that each item fit the Rasch model. The MNSQ for items in the sample ranged from 0.54 to 1.40 (Table 15). Items 1, 4, 7, and 14 had infit MNSQ values above 1.40 and were deleted. Dimensionality results reported above reflect the final model without these items.

Table 15
Use Factor Item Fit Statistics

Calibration Sample				
Item#	Logit Position	SE	Infit MNSQ	Pt-Measure Correlation
2	.41	.23	1.05	.69
3	-.59	.23	1.07	.57
5	-.88	.27	1.13	.53
6	-.16	.20	1.40	.54
8	-.11	.20	1.23	.62
9	1.48	.12	.82	.79
10	2.20	.13	.74	.83
11	-1.67	.40	.92	.57
12	-.87	.34	1.34	.62
13	-1.16	.28	.98	.53
15	1.37	.14	1.00	.73
Validation Sample				
Item#	Logit Position	SE	Infit MNSQ	Pt-Measure Correlation
2	.43	.28	1.25	.62
3	-.20	.23	1.24	.69
5	-1.46	.34	.93	.37
6	-1.06	.26	1.30	.50
8	-.42	.23	1.01	.65
9	1.70	.13	.67	.80
10	2.05	.13	1.15	.74
11	-.94	.37	.54	.76
12	-.65	.33	1.00	.59
13	-.69	.28	.98	.55
15	1.23	.14	1.08	.66

Person fit was also examined to ensure that individuals were answering in a consistent, expected manner. There were three teachers in the calibration sample with MNSQ fit above 4.0 demonstrating they underfit this model; these cases were deleted. Once the model was rerun, all persons fit with MNSQ infit values under 3.67. The validation sample had five teachers that misfit the model and those cases were removed and analysis rerun.

Reliability. Person separation in the calibration sample was 1.11, with reliability of person separation of 0.55, and a Cronbach's alpha of 0.96. Although the Cronbach's alpha is quite high, this is not a desirable spread of persons. Similar findings were seen with the validation sample (Table 14); persons were not sufficiently widely distributed to yield strong reliability.

Scale Use. Results of the Rasch analysis indicated that there were problems with scale use. Category probability curves should demonstrate an even distribution of the categories with clearly advancing steps. Using the initial six categories of Never, Yearly, Quarterly, Monthly, Weekly, and Daily the categories were disordered and very "flat" in their probability curves. Categories were collapsed into Never–Yearly, Quarterly–Monthly, Weekly, and Daily to resolve these issues (Figure 7). Once restructured, Rasch-Andrich thresholds increased with category values with no evidence of step misfit having MNSQ infit values under 2.0 (Linacre, 2015). Teachers used the Weekly and Daily categories more frequently, suggesting our sample was using the assessment process frequently (Table 16). The validation sample's scale use followed the same pattern.

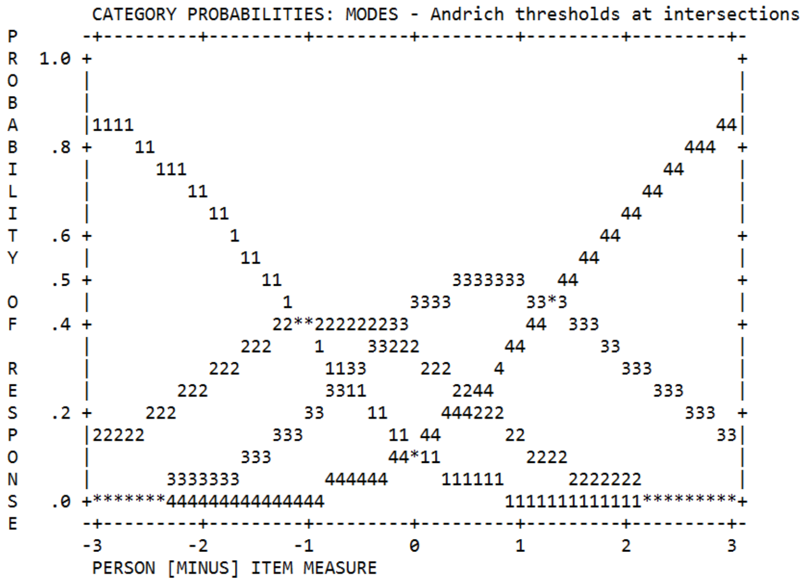


Figure 7. Use Factor Category Probability Curves

*Categories: 1-Never-Yearly, 2-Quarterly-Monthly, 3-Weekly, 4-Daily

Table 16
Use Factor Step Structure

Calibration Sample				
Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
1 Never–Yearly	13%	-1.40	1.06	(-2.51)
2 Quarterly–Monthly	17%	-.33	.93	-.82
3 Weekly	29%	.88	.89	.72
4 Daily	41%	2.08	1.03	(2.67)
Validation Sample				
Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
1 Disagree	12%	-1.18	1.14	(-2.58)
2 Somewhat Agree	18%	-.29	.84	-.81
3 Agree	29%	.89	1.02	.76
4 Strongly Agree	41%	2.31	1.00	(2.66)

Note. Observed percentage is the percent of all responses for that category. Observed average is the average of the measure to produce the responses observed in the category. Infit MNSQ is the average of the infit MNSQs associated with responses in that category. Step Structure is the logit position where the transition is made from a lower to this category.

Targeting and Construct Coverage. The item-person map provided in Figure 8 presents items and persons on the same scale for the calibration sample. Participants are spread evenly on the item-person map, but the majority of the sample falls between 0.5 and 2.0 logit positions. Representation of items and participants along the map suggest this sample of teacher have quite frequent use of the assessment process and that items are too easy to agree with for this sample. The validation sample followed a very similar pattern.

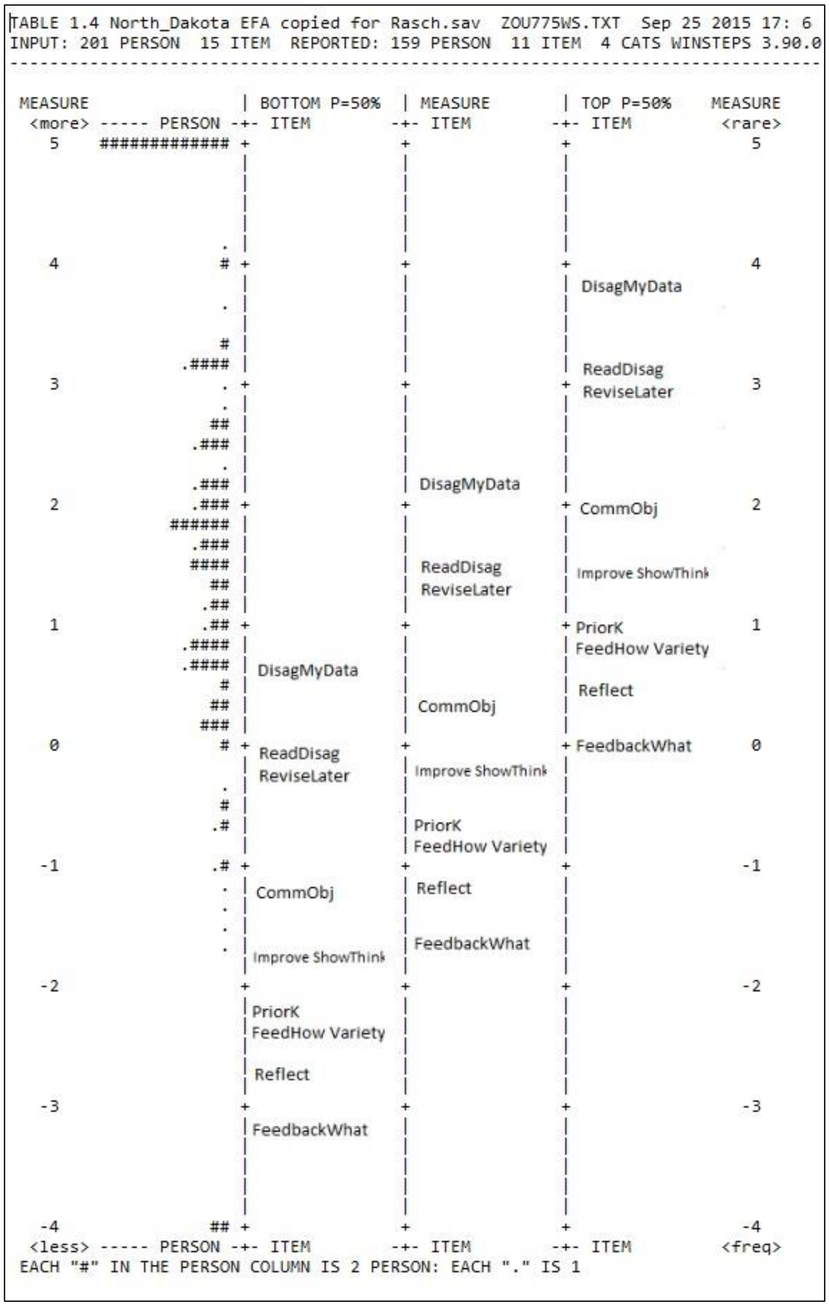


Figure 8. *Use Item-Person Map*

Rasch analysis allows researchers to see how items act on a continuum of "easiest" to agree with to "hardest" to agree with, thus creating a kind of ruler of construct coverage. The items at the top of Figure 8 are questions that were hardest for teachers to agree with, while those at the bottom were the easiest to agree with. This can also be read as those items at the top are done with less frequency and those towards the bottom are done more frequently. The distribution of items, seen on the right-hand side of Figure 2, had item logit values roughly between ±2, indicating the tKUDA is a good measure along this range (Linacre, 2015). Items are spread nicely across persons. There are two areas where the measure could be improved. Items 8 ("improve learning") and 6 ("show thinking") are equal on the "ruler" suggesting that only one of these two items is needed and the second is redundant because it is a measure of the construct at the same location. Item 12 ("giving feedback on how to improve") and item 5 ("using a variety of assessments") had equivalent item positions as well.

Invariance. Differential item functioning (DIF) was examined to ensure that the items were functioning in the same way across demographic groups. DIF is assessed using t-tests of the significance of differences in item logit position. At a significance level of 0.01, there was no DIF in the calibration sample between racial/ethnic identities, sex, or grade level taught for the calibration or validation samples. This means the tKUDA was invariant across those variables. Differences by content taught were seen in the calibration sample for item 11 ("giving feedback on what to improve"), $p < 0.001$ and item 12 ("giving feedback on how to improve"), $p < 0.001$, between elective teachers and all other content groups. Elective teachers report higher use of feedback than science, language arts, social studies, and special education teachers, but lower than math teachers. Differences in content were seen in the validation sample for item 2 ("communicating objectives"), $p < 0.001$, between social studies teachers and all other contents with social studies teachers scoring higher and on item 13 (“reflecting using data”), $p < 0.001$, between elective teachers and all other contents with elective teachers scoring lower. This measure showed some failure of invariance across content area.

Summary. The Use of the Assessment Process measure can be considered unidimensional. Items 1, 4, 7, and 14 showed misfit and were removed. Items were nicely spread throughout the continuum but persons were too clustered together. The measure can be considered reliable with a Cronbach’s alpha of 0.96 and a reliability of item separation of 0.95, but at 0.55 the reliability of person separation could be improved. The original scale of Never, Yearly, Quarterly, Monthly, Weekly, and Daily did not work as intended and was collapsed into four categories of Never–Yearly, Quarterly–Monthly, Weekly, and Daily. Items were spread quite nicely but could be improved by deleting overlapping items and filling in the gaps on the continuum. The measure can be considered invariant across grade level, sex, and race but not across content expertise.

Confirmatory Factor Analysis

Confirmatory factor analysis was used to assess the model seen in the PCA using an independent sample. Analysis was conducted with AMOS 7.0 (Arbuckle, 2006) using raw scores. Full-information maximum likelihood estimation was used to handle missing data. Model fit was assessed using the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA). CFI compares the theoretical model to a null model and is considered sufficient with values of 0.9 and above. RMSEA is sensitive to parsimony of the model with values of 0.1 and below considered acceptable (citation).

The research questions addressed by this analysis was: Is the factor structure confirmed in an independent sample? Here the validation sample was used in the CFA. There were 164 respondents. Demographic information for this sample is seen in Table 17.

Table 17
Validation Sample Participants' Demographics

Race/Ethnicity	Gender	Grade Level	Content Expertise
86.6% White	18.9% Male	17.7% K–2nd	18.3% Math
1.8% Hispanic/Latino	73.2% Female	17.7% 3rd–5th	14.0% Science
0.6% Other		22.0% 6th–8th	25.0% Language Arts
3.0% Prefer not to answer		14.9% 9th–10th	7.3% Social Studies
		24.0% 11th–12th	1.2% Music
			1.8% Art
			3.7% Physical Ed
			4.3% Comp/Tech
			3.7% Foreign Lang
			5.5% Special Ed
			6.7% Other

Knowledge of the Assessment Process. The first CFA model attempted was the one-factor Knowledge model. As seen in Figure 9, the one-factor model consisted of 15 items. The model results show that this was not a good fit to the data, $\chi^2(90) = 362.25$, $p < 0.001$; RMSEA = 0.14; CFI = 0.85. The modification indices indicated misfit with items 9 and 10 which were also seen in the PCA analysis; these items were deleted. The model was rerun with 13 items but the fit statistics did not improve, $\chi^2(65) = 281.19$, $p < 0.001$; RMSEA = 0.14; CFI = 0.87.

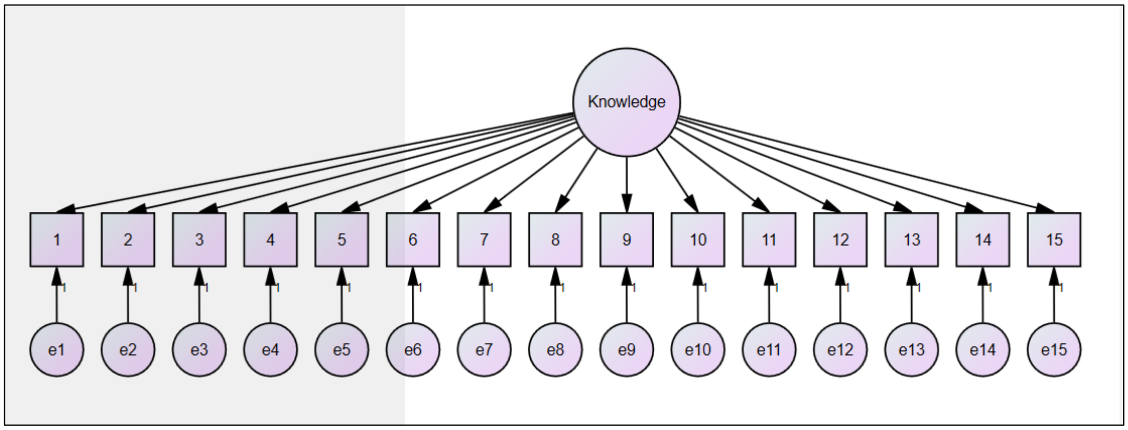


Figure 9. *Knowledge One Factor Model*

The second model attempted was a two factor Knowledge model (Figure 10). Items 9 and 10 comprised a factor and the other 13 items comprised the second factor. This model demonstrated slightly better fit, χ^2 (89) = 330.90, $p < 0.001$; RMSEA = 0.13; CFI = 0.87. Two items, 2 and 12, were removed due to misfit. Item 2 also appeared problematic in the Rasch analysis. The model was rerun with 13 items showing much improved fit χ^2 (64) = 193.40, $p < 0.001$; RMSEA = 0.11; CFI = 0.92. These three fit indices together show that this model had acceptable fit and demonstrated that the two factor model adequately modeled knowledge of the assessment process.

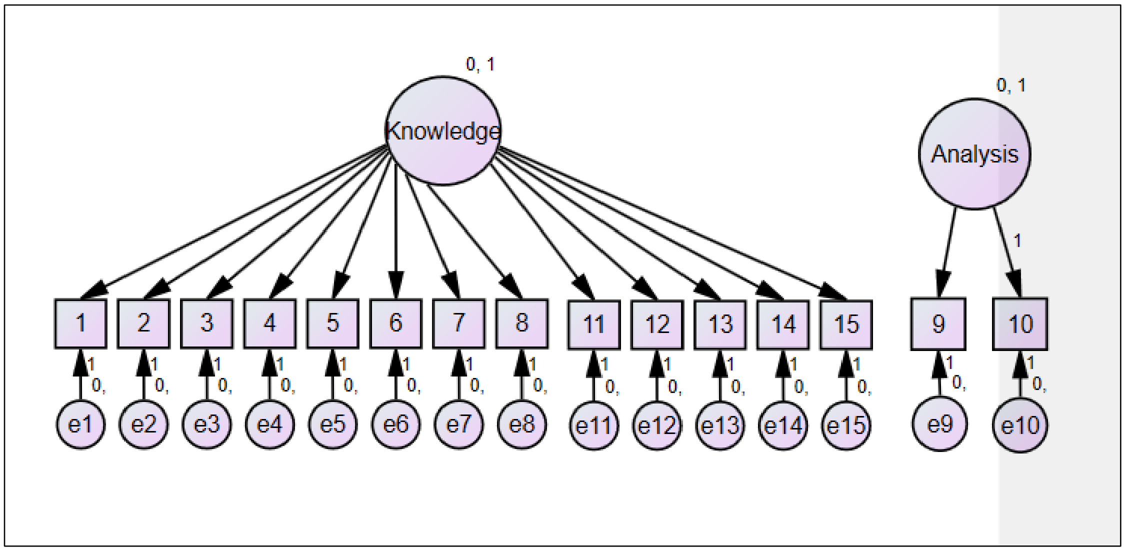


Figure 10. *Knowledge Two Factor Model*

Use of the Assessment Process. The first CFA model attempted was the one factor Use model with all 15 items (Figure 11). This model showed poor fit, χ^2 (90) = 352.17, p < 0.001; RMSEA = 0.13; CFI = 0.68. Items 2 and 15 were removed due to poor fit but the model fit did not improve.

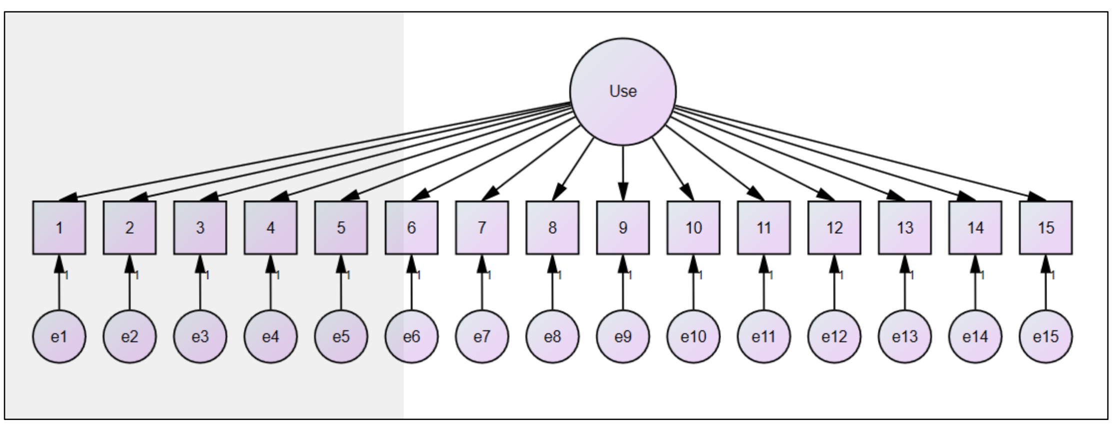


Figure 11. *Use One Factor Model*

Based on the PCA results that showed one large factor and three possibly small factors, a four factor model was tried with Planning Use (items 1–8), Analysis Use (items 9–10), Feedback Use (11–12) and Revision Use (items 13–15) (Figure 12). This model demonstrated much more acceptable fit, χ^2 (88) = 200.20, $p < 0.001$; RMSEA = 0.09; CFI = 0.86. Item 2 still showed misfit and was removed. The model was rerun without this item and demonstrated good fit, χ^2 (75) = 132.79, $p < 0.001$; RMSEA = 0.07; CFI = 0.92.

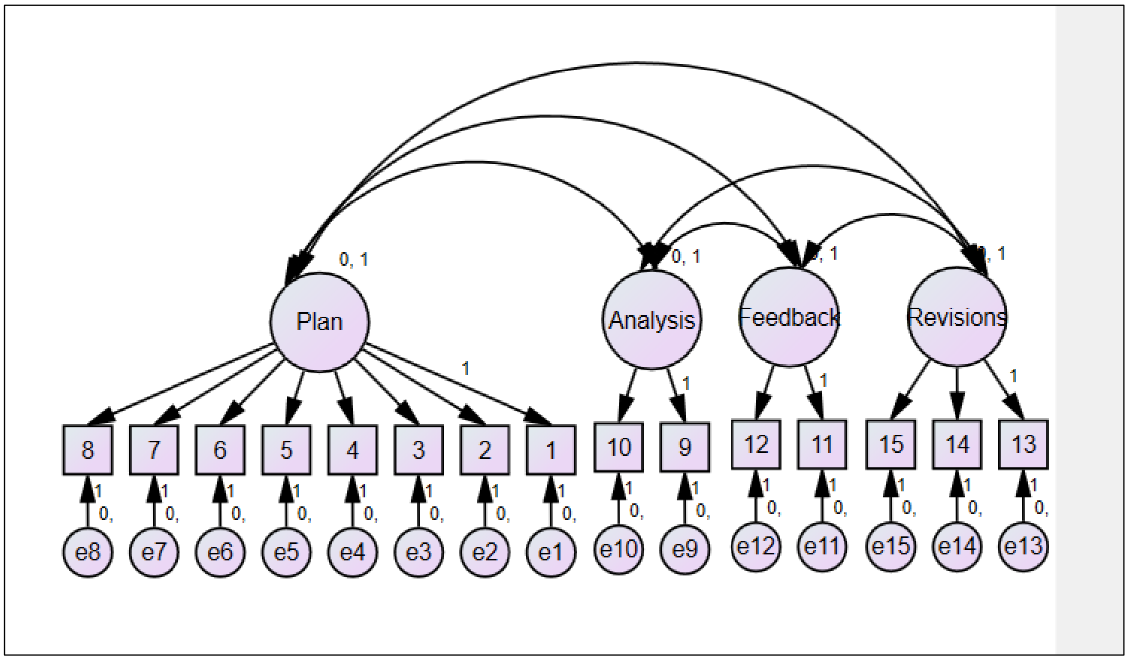


Figure 12. *Use Four Factor Model*

Dimensionality of the tKUDA. Knowledge of the Assessment Process can be considered unidimensional based on PCA, parallel analysis, and Rasch analysis with items 9 (“reading disaggregated data”) and 10 (“disaggregating my own data”) showing misfit. These are two items that seek to understand data analysis strategies and it is not surprising that they act differently when considering a teacher’s knowledge. There are no teacher standards that address these concepts (Association of Childhood Education

International, 2007; National Council for the Social Studies, 2002; National Council of Teachers of English, 2012; National Council of Teachers of Math, 2003; National Science Teacher Association, 2012; Rubric for evaluating Colorado teachers, 2013) and there are several studies that show teachers not doing this (Hoover & Abrams, 2013) or that they have not been adequately taught in teacher preparation programs (Greenberg & Walsh, 2012). Structural equation modeling was used to conduct confirmatory factor analysis (CFA) to compare three models: a single factor with all 15 items, a single factor removing the two items, and a two factor model with the two items as the second factor. The CFA demonstrated a better fitting model using the two factor model. When considering all analyses together there is support for a single construct with a possible smaller facet regarding knowledge of data analysis.

Use of the Assessment Process can also be considered unidimensional, but seems to have four distinct facets to the whole. Principle components analysis showed one dominant factor with an eigenvalue of 5.38 and three very small possible factors with eigenvalues close to 1.0. All items loaded on the unrotated first factor above 0.40, but the rotated matrix shows items falling into four distinct categories of Planning Use (items 1–8), Analysis Use (items 9–10), Feedback Use (11–12) and Revision Use (items 13–15). Parallel analysis gave support for one factor and Rasch also supported a single factor with items 1, 4, 7, and 15 showing misfit. A CFA model with all 15 items was compared to a four factor model. The four factor model demonstrated the best fit. Altogether there is support for a single, overarching construct with four possible smaller facets.

When comparing models, this research relied more strongly on results from Rasch analysis. CFA is based on classical test theory which is focused at the test level and constructed using a linear model where the observed score is equal to the true score plus error. Rasch is based on item response theory which is focused at the item level and constructed using a nonlinear model dedicated by the residuals, $P_{ni}(X) = e^x/(1+e^x)$. CFA is good for exploring multiple dimensions, but Rasch has stricter guidelines on unidimensionality. Wright (1994) notes that misfit in the Rasch model and extreme values of unidimensional factors can be reported as minor factors by factor analysis. Wright also says that if a factor is not confirmed by Rasch analysis then its existence is doubtful. Therefore since Rasch analysis for both factors met unidimensionality requirements, Knowledge and Use were considered single constructs.

Content Validity Analysis

Content validity was explored by correlating the average item difficulty ratings from the expert reviews and cognitive interviews to the item logit position obtained through Rasch analysis. A moderate correlation was found the Knowledge factor (r =0.58). The Use factor showed a strong correlation between experts opinions and Rasch item difficulty positions (r = 0.87).

Convergent Validity Analysis

Convergent validity was explored by adding ten items from the "Using Data to Inform Decisions" survey (Cavalluzzo et al., 2014) to the survey. The internal consistency reliability of those items was $\alpha = 0.87$. Since the tKUDA measure should be measuring the same construct as the Use factor, a correlation of the average Use score and the average of the "Using Data to Inform Decisions" survey items was computed. These two constructs should be strongly correlated to give support for convergent validity. SPSS software was used to perform this correlation analysis.

Of the 301 teachers that responded to the survey, only 254 had complete data for Knowledge items, Use items, and the "Using Data to Inform Decisions" items and were used for this analysis. Demographic information can be seen in Table 18 below.

Table 18
Merged Dataset Participant Demographics

Race/Ethnicity	Gender	Grade Level	Content Expertise
90.6% White	20.8% Male	15.0% K–2nd	19.8% Math
1.5% Hispanic/Latino	79.2%	17.3% 3rd–5th	11.8% Science
1.1% African American	Female	25.9% 6th–8th	22.1% Language Arts
1.5% Other		16.9% 9th–10th	11.4% Social Studies
0.7% Am.Ind/AlaskaNat.		24.8% 11th–12th	3.0% Music
4.5% Prefer not to answer			1.1% Art
			2.7% Physical Ed
			4.6% Comp/Tech
			2.7% Foreign Lang
			10.3% Special Ed
			10.6% Other

A statistically significant relationship was found between Use of the Assessment Process and the Using Data to Inform Decisions scale, r (253) = 0.41, p < 0.001. This is a moderate, positive relationship. This relationship is lower than anticipated, but gaining strong validity for this construct has proved difficult for other studies as well (Merlter & Campbell, 2005; Perry, 2013). The Knowledge of the Assessment Process is a similar, but distinctly different scale and the correlation between Knowledge and the Using Data to Inform Decisions scale was lower, as would be expected, r (255) = 0.20, p = 0.001.

Differences in Subscale Score by Demographic Variables

Analysis of variance (ANOVA) was used to answer research question 2, "Do teachers respond differently to subscales based on demographic differences, such as gender, race/ethnicity, content expertise, and number of years teaching?" Multiple one way ANOVAs were conducted separately to compare each factor, Knowledge and Use, by demographic variables using SPSS software. The significance level was set at 0.05 with no adjustments for type I error. Assumptions for all ANOVAs were tested and met or adjustments made (discussed below). The race/ethnicity variable was collapsed into White/Minority due to small sample sizes and content expertise was collapsed into math, science, social studies, language arts, special education, and elective/other.

There were 301 teachers who responded to the survey. Demographic information about the respondents can be seen in Table 18 above.

No main effect of Knowledge was found for white/minority, F(1, 263) = 0.69, p = 0.41, sex F(1, 261) = 0.75, p = 0.39, content F(5, 255) = 1.29, p = 0.27, or grade level, F(4, 259) = 0.39, p = 0.81. No main effect of Use was found for white/minority, F(1, 261) = 2.03, p = 0.16, sex F(1, 259) = 1.84, p = 0.18, content F(5, 253) = 1.32, p = 0.26, or grade level, F(4, 257) = 0.33, p = 0.86, or years teaching, F(8, 45) = 1.46, p = 0.17.

A main effect of Knowledge was found for Years Teaching, F(3, 252) = 6.17, $p < 0.001$, $\eta^2 = 0.008$. Years of teaching variable was categorized into four groups: 1^{st}–5^{th}

year teachers, 6^{th}–5^{th} year teachers, 16^{th}–25^{th} year teachers, and 26^{th}+ year teachers. The Games Howell post hoc test was used to assess the significance of program differences because homogeneity of variance was violated for this variable. At the $p < 0.05$ level, statistically significant differences were found between 1^{st}–5^{th} year teachers and all other groups, with a lower mean for the less experienced teachers (Table 19).

Table 19

Knowledge by Years Teaching

Years Teaching	Mean	Std. Dev.	N
1^{st}-5^{th} years	2.96	.40	43
6^{th}-15^{th} years	3.30	.47	86
16-25^{th} years	3.18	.39	72
26^{th} + years	3.31	.56	55

Relationships between Factors

Correlation analysis was conducted using a total score for Knowledge and a total score for Use to explore relationships. Both variables met normality assumptions. The significance level was set at 0.05. This relationship was further explored by comparing independent correlations after transforming the r to Fisher's Z. Since there were multiple independent correlations, a chi-square test for independent correlations was used to explore across groupings.

Research questions addressed by these analyses were: Is there a relationship between Knowledge and Use? Does this relationship differ by demographic variables such as number of years teaching and content expertise?

A statistically significant relationship was found between teachers' Knowledge and Use of the assessment process, r (300) = 0.47, $p < 0.001$. This is a moderate, positive relationship meaning as Knowledge increases, Use also increases. Squaring the correlation revealed that these two variables have 22.1% of their variance in common; 22% of Use is explained by Knowledge.

To examine if this relationship differed by background variables of Content Expertise and Years Teaching, a chi-square test for independent correlations was used to explore differences in correlations across these groups. Years teaching was grouped into the following categories: 1–5 years, 6–15 years, 16–25 years, 26+ years. Content expertise was compared between math, science, language arts, social studies, special education, and elective/other teachers. Although differences in relationships between Content Expertise and Years Teaching are noticeable (Table 20), no significant overall differences in correlations were found for Content Expertise, χ^2 (5) = 5.18, p = 0.39 or for Years Teaching, χ^2 (3) = 1.05, p = 0.80. To further verify, an independent correlation difference test between two correlations was conducted and found a significant difference between correlations for language arts and special education, z = 2.06, p = 0.04. It is important to note that the small sample size per group affects significance; a future, larger sample size may yield significant results.

Table 20

Knowledge-Use Correlations By Group

Content Expertise			
Content	*r*	*p*	*n*
Math	.37	.007	52
Science	.28	.14	30
Language Arts	.60	<.001	58
Social Studies	.61	<.001	30
Special Education	.18	.37	26
Elective/Other	.40	.001	63
Years Teaching			
Years	*r*	*p*	*n*
1st-5th year teachers	.48	.001	42
6th-15th year teachers	.51	<.001	86
16th-25th year teachers	.36	.002	72
26+ year teachers	.39	.004	54

Descriptive Statistics for Practical Applications

To answer the fourth research question, "What are the practical applications of data and assessments that teachers rely on and what data sources are they using?" descriptive statistics were utilized. Descriptive statistics report how teachers use data in their classroom for practical applications such as informing instruction or giving parent feedback. Descriptions focus on what sources of data teachers' rely on in order to perform these tasks and then looks at each type of data and what teachers use it for. Analysis was conducted using SPSS software.

The research question addressed by this analysis is: What are the practical applications of data and assessments that teachers reply on and what data sources are they using?

Teachers were asked to respond to 22 questions, or categories, of practical applications of data based on what type of assessment they used to do this. For example, the question “I inform my instruction using…” was followed by choices of: planned formative assessments, in the moment formative assessments, classroom summative assessments, school-level assessments, district-level assessments, state-level assessments, or I never do this with data. Teachers could check one or multiple assessments types. For analysis these were coded 1 = yes, and 0 = no. All categories were converted to percentage of teachers reporting use of that assessment for that practical application (Table 21).

Table 21

Practical Applications of Data and Assessments in Percentage of Agreement

Practical Application	Planned formative assess	Immediate formative assess	Classroom summative assess	School-Level assess	District-Level assess	State-level assess	Never w/ data
I inform (drive) my instruction	48.1	42.4	50.0	24.9	20.3	18.9	0.8
I set new learning goals	34.9	25.9	45.4	21.4	17.0	16.5	0.8
I differentiate instruction	44.9	42.7	45.1	16.2	12.2	10.0	0.3
I create student groups	36.5	33.0	38.9	15.9	14.6	9.2	5.7
I reteach or review content	42.2	42.7	50.8	14.6	8.4	6.2	0.5
I identify gaps in learning or target skills	40.5	36.2	49.7	21.4	15.1	13.5	1.1
I reflect on and revise instruction	41.9	42.4	51.1	18.9	12.2	11.9	0.8
I gauge my students' engagement	43.0	50.5	27.3	6.2	3.0	2.2	0.8
I get a "feeling for" incoming students	34.3	34.6	25.4	17.3	20.3	20.0	2.7
I give parents feedback of student learning	30.5	20.3	57.3	30.0	23.0	22.4	2.4

I give students feedback on their learning	46.8	41.4	58.1	26.8	17.6	15.4	0.3
I facilitate student goal setting	30.8	21.4	41.6	18.1	17.0	13.2	7.8
I identify student strengths and weaknesses	48.1	43.2	53.8	24.9	19.2	18.9	0.5
I explore patterns across students	33.8	27.6	46.5	22.7	19.5	19.2	3.5
I demonstrate/ prove student achievement	33.0	23.2	55.7	29.7	23.5	20.5	0.5
I track growth/progress monitoring	36.2	25.9	51.4	27.0	21.9	15.9	1.4
I monitor target students	42.7	34.1	50.3	20.5	17.3	10.0	2.7
I disaggregate content	28.1	20.3	46.5	23.8	20.8	16.5	4.1
I disaggregate demographics	16.2	11.4	23.5	20.5	22.4	23.5	17.0
I decide or recommend student placement	27.3	20.8	45.1	32.4	27.3	23.5	8.6
I predict students' future scores	26.2	21.1	33.5	18.9	18.4	14.9	15.1
I triangulate learning	30.3	25.4	40.8	21.4	18.1	15.9	16.2

Given the 22 categories of practical applications, teachers report using some source of data for all applications (Table 21). This does not mean all teachers use data for all categories, but no teachers reported “never with data” across all categories. Teachers rely heavily on data to inform instruction, differentiate instruction, reteach and review content, reflect on and revise instruction, give students feedback, and to identify students’ strengths and weaknesses; these categories have the highest percentages. Even so, sources of data were relied on differently depending on application and different types of assessments were used for different reasons. While this is the same concept, it is easier to consider these separately, essentially considering across rows first and then columns.

First, to consider what types of data teachers are relying on to do applications, I looked across the rows in Table 21. Categories where teacher reported use at 40% or above were included. This is a subjective cut-off intended to help get a “feeling for” what teachers are using.

Teachers seem to be relying on their own assessments (both formative and summative assessments) to identify gaps and target skills (40.5% to 49.7%), monitor target students (42.7% to 50.3%), inform instruction (42.4% to 50.0%), differentiate instruction (42.7% to 45.1%), reteach and review content (42.2% to 50.8%), reflect on and revise instruction (41.9% to 51.1%), give students feedback (41.4% to 58.1%), and identify student strengths and weakness (43.2% to 53.8%). Teachers specifically rely on their own classroom summative assessments to set new learning goals (45.4%), give parents feedback (57.3%), facilitate student goal setting (41.6%), explore patterns across students (46.5%), demonstrate/prove student achievement (55.7%), track growth/progress monitor (51.4%), break down data by content (46.5%), decide/recommend student placement (45.1%), and triangulate student learning (40.8%). Regarding only formative assessments, teachers rely on these specifically to gauge student engagement (50.5%).

Interestingly, teachers are *not* heavily relying on data from any assessment to get a feeling for incoming students (under 34.6%), create student groups (under 38.9%), break down data by demographics (under 23.5%), and predict students' future scores (under 33.5%).

Second, to distinctly explore what each type of assessment is used for, I examined the columns in Table 21. The cut-off for a practical application to be included was 40% or higher was, but for the formal data sources this had to be dropped to 20%. These are also subjective cut-offs to get a "feeling for" what teachers are using each data source to do in their classrooms.

Teachers use planned formative assessments to inform instruction (48.1%), differentiate instruction (44.9%), reteach and review content (42.2%), identify gaps and target skills (40.5%), reflect on and revise instruction (41.9%), give students feedback (46.8%), identify student strengths and weakness (48.1%), track growth/progress monitor (42.7%), and gauge student engagement (43.0%).

Teachers use in the moment formative assessments to inform instruction (42.4%), differentiate instruction (42.7%), reteach and review content (42.7%), reflect on and revise instruction (42.4%), give students feedback (41.4%), gauge student engagement (50.5%), and identify students' strengths and weaknesses (43.2%).

Teachers rely heavily on their own classroom summative assessments. These assessments are used to inform instruction (50.0%), set new learning goals (45.5%), differentiate instruction (45.1%), reteach and review content (50.8%), identify gaps and target skills (49.7%), reflect on and revise instruction (51.1%), give parents feedback (57.3%), give students feedback (58.1%), facilitate student goal setting (41.6%), identify student strengths and weakness (53.8%), explore patterns across students (46.5%), demonstrate/prove student achievement (55.7%), track growth/progress monitor (51.4%), monitor target students (50.3%), break down data by content (46.5%), decide/recommend student placement (45.1%), and triangulate learning (40.8%).

Teachers did not report strong practical applications of the formal assessments: school-level assessments, district-level assessments, and state-level assessments. Roughly 40% of the categories were *not* reported being used by 80% of teachers. The following categories were reported as practical applications of formal data that at least 20% of teachers used to inform instruction (school-level 24.9%, district-level 20.3%), set new learning goals (school-level 21.4%), identify gaps and target skills (school-level 21.4%), get a feeling for incoming students (district-level 20.3%, state-level 20.0%), give parents feedback (school-level 30.0%, district-level 23.0%, state-level 22.4%), give students feedback (school-level 26.8%), identify students' strengths and weaknesses (school-level 24.9%), explore patterns across students (school-level 22.7%), demonstrate/prove student achievement (school-level 29.7%, district-level 23.5%, state-level 20.5%), track growth/progress monitor (school-level 27.0%, district-level 21.9%), monitor target students (school-level 20.5%), break down data by content (school-level

23.8%, district-level 20.8%), break down data by demographics (school-level 20.5%, district-level 22.4%, state-level 23.5%), decide/recommend student placement (school-level 32.4%, district-level 27.3%, state-level 23.5%), and triangulate learning (school-level 21.4%).

Since the percentage of teachers report doing specific applications using specific assessment types, Spearman's rank order correlations were conducted to explore this further. Items within assessment type were rank ordered to see if this ordering was consistent across assessment types. These correlations are seen in Table 22. Strong positive relationships were found between the two formative assessments ($r = 0.92$, $p < 0.001$), and between school-level summative assessments and classroom summative assessments ($r = 0.60$, $p = 0.003$), and between all the formal assessments (school-level, district-level, and state-level; Table 22). This means teachers report using these assessments with similar relative frequency. Moderate to strong negative relationships were found between district-level assessments and both types of formative assessment (immediate and planned for; Table 22) and between state-level assessments and both types of formative assessments (Table 22). This demonstrates that teachers report using these types of assessments in opposite manners. Additionally, strong negative relationships were found between never using data and teachers' own assessments (immediate formative, planned formative, and classroom summative; Table 22). This gives evidence that teachers who tend to not use data are more likely to use data from their own assessments.

Table 22
Spearman Rank Order Correlation of Practical Application

	PlanForm	ImmForm	ClassSum	School	District	State	Never
PlanForm	1						
ImmForm	.92***	1					
ClassSum	.41	.20	1				
School	-.17	-.40	.60**	1			
District	-.52**	-.66***	.16	.77***	1		
State	-.48*	-.57**	.08	.72***	.91***	1	
Never	-.78***	-.71***	-.65***	-.05	.31	.25	1

*0.05
** 0.01
*** 0.001

CHAPTER 4: DISCUSSION

Teachers' classroom practices have not been adequately considered when policies regarding assessments are put into place (Jennings, 2012), which leads to assessments becoming an "extra thing" teachers must do. In reality, teachers assess their students daily, even minute by minute, using formative assessment techniques (Williams, 2011), which guides student learning towards a teacher's chosen summative assessment (Wiggins & McTighe, 2005). Assessment is not a "thing" that teachers do, but rather a process that teachers engage in (Heritage, 2007). In order to capture these classroom practices around data and assessments, this study focused on K–12 teachers' knowledge and use of the assessment process.

Before the teachers Knowledge and Use of Data and Assessments (tKUDA), the measures of teachers' use of data and assessments had questionable reliability or were limited in their scope. The Assessment Literacy Inventory (Mertler & Campbell, 2005) and the modified Classroom Assessment Literacy Inventory (Perry, 2013) are like a test for teachers where a scenario is given and teachers choose the best answer. These had low reliabilities for current teachers (0.54) and moderate reliabilities for pre-service teachers (0.74). Lysaght and O'Leary (2013) created a measure how teachers use formative assessments and Cavalluzzo et al. (2014) created a more comprehensive survey to explore math teachers' knowledge and use of assessments, but these are only a piece of teachers' actual classroom practice. The tKUDA was designed to capture K–12 teachers' knowledge and use of the assessment process and practical applications of data in order to understand current classroom practice, compare practice and policy to identify gaps, and match professional development to meet these gaps. This chapter summarizes the results and implications of the tKUDA.

Study Overview

The purpose of this measure development study was two-fold: to create an instrument that can measure teachers' Knowledge and Use of data and assessment and then to explore differences between respondents and relationships between factors. There were four main research questions for this study:

1. What is the measured construct?
    a. Do items factor appropriately into distinct domains of Knowledge and Use? Is the factor structure confirmed in an independent sample?
    b. Is there adequate reliability and validity for each of these factors?
    c. Are the response scales used appropriately?
    d. Is the measure well targeted?
    e. Which items are the hardest and which are easiest for the teachers to agree with?
2. Do teachers respond differently to subscales based on demographic differences, such as gender, race/ethnicity, content expertise, and number of years teaching?

3. Is there a relationship between Knowledge and Use? Does this relationship differ by demographic differences such as number of years teaching and content expertise?
4. What are the practical applications of data and assessments that teachers rely on and what data sources are they using?

The following sections break down and summarize each research question with major findings and importance. Implications, limitations, recommendations for future research, and the final tKUDA measure are also given in separate sections below.

Question 1: What is the measured construct?

The first research question sought to understand the construct of the tKUDA. Multiple sub-questions were needed to explore this question, including dimensionality, reliability and validity, scale use, and item difficultly.

Do items factor appropriately into distinct domains of Knowledge and Use? When considering the 15 Knowledge items and the 15 Use items, there are two distinct factors with items falling into the appropriate category (Appendix E). Knowledge of the assessment process can be considered unidimensional, but items regarding analysis show misfit. This is not surprising as there are no teacher standards that address these concepts (Association of Childhood Education International, 2007; National Council for the Social Studies, 2002; National Council of Teachers of English, 2012; National Council of Teachers of Math, 2003; National Science Teacher Association, 2012; Rubric for evaluating Colorado teachers, 2013), and there are several studies that show teachers not doing this (Hoover & Abrams, 2013) or have not been adequately taught in teacher preparation programs (Greenberg & Walsh, 2012). Use of the assessment process can also be considered unidimensional, but seems to have four distinct facets to the whole: planning use (items 1–8), analysis use (items 9–10), feedback use (11–12), and revision use (items 13–15). Items 1, 4, 7, and 15 demonstrated misfit and are candidates for replacement.

Is the factor structure confirmed in an independent sample? Analysis was conducted using two independent samples. Principle components analysis and parallel analysis were conducted with the calibration sample and confirmatory factor analysis using structural equation modeling employed the validation sample. Rasch analysis also used these two samples to run two separate models which complemented each other with no unexpected or drastically different results. However, different items showed misfit in PCA, CFA, and Rasch analysis. PCA showed no concerns with Knowledge items but problems were seen with items 14 and 15 for Use. CFA showed items 2 and 12 as misfit for Knowledge and item 2 misfit for Use. Rasch showed misfit for items 2, 9, and 10 for Knowledge and items 1, 4, 7, and 14 for Use. This shows that the Knowledge factor structure is confirmed with an independent sample but item 2 is a candidate for replacement. The Use factor shows a few items being problematic for one sample and different items having misfit with the other sample. As a whole the Use factor structure is working in the same manner, but items 1, 2, 4, 7, 14, and 15 need to be explored in further depth with another sample.

Is there adequate reliability and validity for each of these factors? Support was found for reliability of the tKUDA. Cronbach's alpha for the Knowledge factor was

0.95 and 0.96 for the Use factor. Additionally, strong reliability was seen using person separation in the Rasch model for the Knowledge factor, but not for the Use factor.

Content validity was accomplished through the opinions and modifications gained during expert and cognitive interviews to verify that the measure was actually determining teachers' Knowledge and Use of data and assessments (DeVellis, 2003). A moderate correlation was found between expert ratings of item difficultly and item logit position for the Knowledge factor and a strong positive correlation was found for the Use factor.

Convergent validity explored thought correlations to the Using Data to Inform Decisions survey (Cavalluzzo et al., 2014) to the field administration survey. A moderate, positive relationship was found between this survey and the Use factor. While this is lower than anticipated, gaining strong validity for this construct has proved difficult for other studies as well (Merlter & Campbell, 2005; Perry, 2013).

Credibility, a qualitative method of demonstrating validity, was used in order to further explore convergent validity (Creswell, 2007). This was established using the pilot study to understand if the survey results matched teacher practice. There was robust agreement between the professional developer at the school and the tKUDA survey results which helped demonstrate that the tKUDA is a valid tool to be used to help guide professional development.

Construct validity was demonstrated through Rasch analysis to examine ratios between categories, test scale use, and explore category structure and function (DeVellis, 2003). Both original scales were modified as discussed below. Both scales used all categories with each category used as expected which demonstrates the final scale decisions for the tKUDA are valid.

Internal structure validity was assessed using PCA and Rasch (DeVellis, 2003) to understand item-factor relationships and differential item functioning (DIF) to test if items were answered differently across demographic groups. All items factored into the expected, appropriate construct (Appendix E). DIF showed that the tKUDA can be considered invariant across race, sex, and grade level as no consistent differences were seen for these variables, which means items are being answered in the same manner regardless of these demographics. Differences are seen between years teaching and content expertise, meaning the responses of these items changes based on these demographics. Specific information on these findings are discussed below under research question two.

Are the response scales used appropriately? The items for the Knowledge scale were on a 1–4 scale from strongly disagree to strongly agree. This scale was modified based on results seen in the pilot sample to disagree, somewhat agree, agree, and strongly agree. Rasch analysis showed that this scale works very well for these items with all categories being used in expected and appropriate ways (Figure 5, Table 13).

The items for the Use scale were on a 1–6 scale of never, yearly, quarterly, monthly, weekly, and daily. Rasch analysis showed problems with category use, step structure, and probability curves. The scale was collapsed to never to yearly, quarterly to monthly, weekly, and daily. Once categories were collapsed, all scale statistics fell into appropriate ranges (Figure 7, Table 16).

Is the measure well targeted? Rasch analysis puts items and persons on the same continuum creating a "ruler" so we can understand how persons are spread throughout the items. The Knowledge factor is well targeted for both samples of teachers (Figure 6). Person spread is very wide on this scale with a large majority of teachers falling on the positive side, meaning they report having strong agreement in their knowledge of the assessment process. More items are needed to capture higher levels of knowledge; items that would be harder for teachers to agree with. Items 7 ("prove learning") and 6 ("show thinking") are equal on the "ruler" suggesting that only one of these two items is needed and the second is redundant, as it is a measure of the construct at the same location. Item 4 ("choose assessments") and item 15 ("revise later") had equivalent item positions as well. Finally there were four items that fell together: item 3 ("assessing prior knowledge"), item 13 ("reflection"), item 5 ("using a variety of assessments"), and item 11 ("giving feedback on what to improve"). These would be items to consider replacing with harder items.

The Use factor can also be considered well targeted. Here the item spread is more ideal and spread throughout persons better than the Knowledge factor (Figure 8). Persons are also spread out toward the positive end, meaning teachers tend to report higher frequencies of use of the assessment process. There are not many overlapping items on the continuum but larger gaps exist that could be filled. Items 8 ("improve learning") and 6 ("show thinking") are equal on the "ruler" suggesting that only one of these two items is needed and the second is redundant, as it is a measure of the construct at the same location. Item 12 ("giving feedback on how to improve") and item 5 ("using a variety of assessments") had equivalent item positions as well. These are items to consider replacing in order to fill in the gaps in the ruler.

Which items are the hardest and which are easiest for the teachers to agree with? Rasch analysis allows researchers to see how items act on a continuum of "easiest" to agree with to "hardest" to agree with, thus creating a ruler of construct coverage. This is important to note for those working with pre-service teachers and providing professional development opportunities. Knowledge of the Assessment Process items are seen in order from hardest to easiest in Table 23, and Use of the Assessment Process items are seen in Table 24.

Table 23

Knowledge Factor Item Difficulty Rank

Hardest item to agree with ↑	I know how to evaluate evidence from assessments in order to prove student learning.
	I know how to effectively use assessments to show students' *thinking*, not just their answers.
	I know how to evaluate evidence from assessments in order to improve student learning.
	I know how to give students specific feedback on how to improve.
	I know how to revise my instructional practices immediately (on the fly) based on evidence from assessments (observations, questioning, quizzes, tests, projects, etc.).
	I know how to assess my students for prior knowledge.
	I know how to use a variety of assessment techniques.
	I know how to give students specific feedback on what they need to improve.
	I know how to reflect on my instructional practices based on evidence from my assessment techniques (observations, questioning, quizzes, tests, projects, etc.).
↓ Easiest item to agree with	I know how to choose or create an assessment strategy that will measure my specific learning objective (like a task, project, discussion, exit slip, quiz, test, etc.).
	I know how to revise my instructional practices for the next year based on evidence from assessments (observations, questioning, quizzes, tests, projects, etc.).
	I know how to set specific learning goals/objectives.

Table 24

Use Factor Item Difficulty Rank

Hardest item to agree with ↑	I break down my own assessments based on gender, race/ethnicity, IEP, GT, content standard, etc.…
	I *typically* read data that is broken down for me based on gender, race/ethnicity, IEP, GT, content standard, etc.…
	I revise my instructional practices for the next year based evidence from assessment techniques (observations, questioning, quizzes, tests, projects, etc.)…
	I *typically* communicate my learning goals to students using multiple methods (posting in the room, via PowerPoint, verbally, having students write it, having students discuss it, reviewing it, highlighting it during lesson, checking for understanding specifically based on objective, etc.)…
	I evaluate evidence from assessments in order to improve student learning…
	I effectively use assessments to show students' *thinking*, not just their answers…
	I pre-assess my students…
	I use a *variety* of assessment techniques…
↓ Easiest item to agree with	I give students specific feedback on how to improve…
	I reflect on my instructional practices based on evidence from assessment techniques (observations, questioning, quizzes, tests, projects, etc.)…
	I give students specific feedback on what they need to improve…

It is interesting to note that item order varied between knowledge and use. Just because a teacher reported high knowledge of an item does not mean she reported high frequency of doing said item. For example, teachers had a harder time agreeing with knowing how to use assessments to show student thinking and to improve learning, but these two items fall in the middle of difficulty on the Use scale; teachers seem to struggle with their knowledge of these items but they do them fairly frequently. On a positive side, teacher report reflecting on practice using data as "easy" and do this quite frequently. Conversely, knowing how to revise instruction for the next year was easy to agree with but done less frequently as would be expected.

Question 2: Do teachers respond differently to subscales based on demographic differences, such as gender, race/ethnicity, content expertise, and number of years teaching?

The tKUDA can be considered invariant across race, grade level, and sex. Differences are seen between Knowledge and number of years teaching (coded: 1st–5th year teachers, 6th–15th year teachers, 16th–25th year teachers, and 26th+ year teachers) with evidence for stronger confidence in knowledge for more experienced teachers. Differences are also seen on an item level between Use and content expertise (coded: math, science, language arts, social studies, special education, and electives). Elective teachers report higher use of feedback than science, language arts, social studies, and special education teachers, but lower than math teachers. Elective teachers also responded lower than all other contents on communicating objectives.

Question 3: Is there a relationship between Knowledge and Use?

A moderate, positive relationship was found between Knowledge and Use. This means that as Knowledge increases, Use also increases, however only 22% of Use is explained by Knowledge. While it is not surprising to find a positive relationship between these two variables, it is surprising that the correlation is only moderate. It is easy to assume that if teachers know how to do something, they will actually do this in practice. This gives empirical evidence to support the flaws of that assumption. There are many factors that influence how a teacher will use data and assessments (Datnow et al., 2012); knowledge is only explaining a small piece of this puzzle.

Does this relationship differ by demographic differences such as number of years teaching and content expertise? To explore if this relationship differs for teachers in different content areas, the data was divided by content expertise and years of teaching with separate correlations conducted for each group (Table 20). No significant difference was found for years teaching but a significant difference was seen between content expertise, specifically between language arts and special education teachers. It is important to note that the small sample size per group affects significance; a future, larger sample size may yield additional significant results. Differences by content were seen by average Use score using ANOVA and this hints that differing relationships by content are also quite possible. Further analysis is needed to verify this finding.

Question 4: What are the practical applications of data and assessments that teachers rely on and what data sources are they using?

Teachers were asked to respond to 22 questions, or categories, of practical applications using data based on what type of assessment they used to do this. For example, the question "I inform my instruction using…" was followed by choices of: planned formative assessments, in the moment formative assessments, classroom summative assessments, school-level assessments, district-level assessments, state-level assessments, or I never do this with data. Teachers could check one or multiple assessments types.

Given the 22 categories of practical applications, teachers report using some source of data for all applications (Table 21). This does not mean all teachers use data for all categories, but no teacher reported "never with data" across all categories. Teacher rely heavily on data to inform instruction, differentiate instruction, reteach and review content, reflect on and revise instruction, give students feedback, and to identify students' strengths and weaknesses; these categories have the highest percentages. Even so, sources of data that teachers relied on differed depending on the application (category) and different types of assessments are used for different reasons.

Teachers seem to be relying on their own assessments (both formative and summative assessments) to identify gaps and target skills, monitor target students, inform instruction, differentiate instruction, reteach and review content, reflect on and revise instruction, give students feedback, and identify students' strengths and weaknesses. Teachers rely specifically on their own classroom summative assessments to set new learning goals, give parents feedback, facilitate student goal setting, explore patterns across students, demonstrate/prove student achievement, track growth/progress monitor, break down data by content, decide/recommend student placement, and triangulate student learning. Teachers specifically rely on their own formative assessments to gauge student engagement. Interestingly, teachers are *not* relying heavily on data from any assessment to get a feeling for incoming students, create student groups, break down data by demographics, and predict students' future scores.

Another way to consider the same information is by looking specifically at the data source and what it is being used for. Teachers use planned formative assessments to inform instruction, differentiate instruction, reteach and review content, identify gaps and target skills, reflect on and revise instruction, give students feedback, identify students' strengths and weaknesses, track growth/progress monitor, and gauge student engagement. Teachers use in the moment formative assessments to inform instruction, differentiate instruction, reteach and review content, reflect on and revise instruction, give students feedback, gauge student engagement, and identify students' strengths and weaknesses. Teachers rely heavily on their own classroom summative assessments using these to inform instruction, set new learning goals, differentiate instruction, reteach and review content, identify gaps and target skills, reflect on and revise instruction, giving parents feedback, giving students feedback, facilitating student goal setting, identifying students' strengths and weaknesses, exploring patterns across students, demonstrate/prove student achievement, track growth/progress monitor, monitor target students, break down data by content, decide/recommend student placement, and triangulate learning.

Teachers did not report strong practical applications of formal assessments (school-level assessments, district-level assessments, and state-level assessments). In fact, roughly 40% of the categories were *not* reported being used by 80% of teachers. The following categories were conveyed as practical applications of formal data that only 20–35% of teachers reportedly use: inform instruction, set new learning goals, identify gaps and target skills, get a feeling for incoming students, giving parents feedback, giving students feedback, identifying students' strengths and weaknesses, exploring patterns across students, demonstrate/prove student achievement, track growth/progress monitor, monitor target students, break down data by content, break down data by demographics, decide/recommend student placement, and triangulate learning. This finding is of particular interest in that the policy makers and administration place high value on these types of assessment (Shen & Cooley, 2008). This gives further evidence that formal assessments are disconnected from classroom practice.

Finally, since the percentages varied between undertaking specific applications using specific assessment types, Spearman's rank order correlations were conducted to explore this further. Items within assessment type were rank ordered to see if this ordering was statistically significant. These correlations are seen in Table 22. Strong positive relationships were found between the two formative assessments, between school-level summative assessments and classroom summative assessments, and between all three formal assessments. This means teachers report using these assessments in the same manner with similar relative frequency; similar applications are being used. Moderate to strong negative relationships were found between district-level assessments and both types of formative assessment and between state-level assessments and both types of formative assessments. This demonstrates that teachers report using these types of assessments in opposite manners. Additionally, strong negative relationships were found between never using data and teachers' own assessments (both formative and summative), which gives evidence that teachers who tend to not use data are at least more likely to use data from their own assessments.

Implications for Professional Development and Teacher Preparation Programs

There are several implications of the tKUDA that were seen from this study. First, this measure is intended to be used to match policy and practice and the pilot gives evidence of the tKUDA's effectiveness in doing so. The measure also allows for and supports differentiation in teacher trainings. The results give quantitative support that there is more to how teachers use data and assessments than just understanding their level of knowledge of data and assessments. The concerns with analysis items demonstrate the gap between assessment literacy and informing instruction. Finally, the sources of data that teachers rely on changes based on the practical classroom applications with teachers relying more on their own various assessments instead of formal assessments.

The tKUDA was designed to be a tool to gauge current classroom practice regarding data and assessment, specifically focused on the assessment process. The intended use is for professional developers and teacher preparation programs to understand current practice and compare this to the intended practice based on their goals or policies. The tKUDA can show strengths of teachers and needs of teachers once these goals or policies are placed side-by-side with the tKUDA results. This allows for the celebration of successful teachings and trainings and for the identification of gaps to be targeted for future training sessions. No position is taken by the author regarding tKUDA results, as what the important aspects of the assessment process are varies by administrators' opinions and school needs. This measure is not intended to be a neutral, stand-alone data point; instead it is to be used in conjunction with administrators' objectives to pin-point professional development needs. It is also not intended to be used to evaluate teachers, although it could be used as a measure administered pre- and post-treatment to gauge professional development effectiveness.

Another implication is for differentiation, which is a standard part of a teacher's practice (Good, 2007), but is not always used in teacher trainings (MacFarlane, 2012). This study gives evidence for possible differentiation considerations during professional development. Differences were noticed between teachers' knowledge and the number of years they had been teaching, specifically between the newer teachers (years 1–5) and the older teachers (all groups above 6 years), with newer teachers reporting less confidence in their knowledge. While this is not a surprising finding, it gives further evidence that newer teachers may need more support in specific ways. Additional evidence was found for differences between use and content expertise. Conducting trainings by content expertise could yield greater, more meaningful trainings and changes in teacher practice.

Another finding of interest is the fairly low correlation between Knowledge and Use. While a correlation of 0.47 can be considered moderate, this only explains 22% of the variance; Knowledge only explains 22% of Use. When conducting teacher trainings, it is easy to believe in the assumption, "I taught it; now they will do it." This could make the professional developer frustrated when these trainings are not put into teachers' practice. This study gives evidence that it takes more than just Knowledge to influence Use. In fact around 78% of Use is explained by something other than Knowledge. Datnow et al. (2012) confirm that the process of data use by teachers is complex, multilayered, and influenced by teacher interpretations and even by social interactions.

The two items on analysis were problems in the Knowledge factor and the hardest items to agree with on the Use factor. This supports many other studies that show teachers lack an understanding of data analysis (Taylor, 2009) or just are not doing this in practice (Hoover & Abrams, 2013). While teachers do not need to be data analysts, they do need basic skills to understand and interpret data (Greenberg & Walsh, 2012; Taylor, 2009; Wong & Lam, 2007). We expect teachers to know types of assessments, what they are for, choose correct assessments to match learning goals, design assessments, and give assessments (Shen & Cooley, 2008; Siegel & Wissehr, 2011; Stronge & Grant, 2009). We also expect teachers to give feedback to parents, students, and the administration based on their data, reflect on and revise their practice, and make instructional decisions based on information from assessments (Brookhart, 2011; Greenberg & Walsh, 2012; Shepherd, et al., 2011). Analysis is the bridge between these activities and at least some basic analysis skills are needed (Greenberg & Walsh, 2012; Taylor, 2009). It is not surprising that teacher understanding and skill are varied and lacking in analysis as there are no teacher standards that cover this. Expectations of data analysis from policy-makers, administration, and university teacher preparation programs vary (Greenberg & Walsh, 2012; Marsh, et al., 2009; Wayman & Stringfield, 2006). The tKUDA is not designed to measure or monitor these expectations, instead it attempts to capture classroom practices that can be compared to these varying perceptions. This is why the analysis items were kept in the measure with a suggestion that the wording could be changed to better capture specific expectations.

Not surprisingly, teachers reported using their own formative and summative assessments with much higher frequency than formal assessments. This is seen in other studies as well (Datnow, 2012; Heritage, 2007; Hoover & Abrams, 2013; Shen & Cooley, 2008; Taylor 2009). Use of formative assessments and strategies to incorporate formative assessments have jointly been a strong focus in teacher education in the last several years (Heritage, 2007; Williams, 2011), so the fact that teachers rely on these is a success story. In this study, the high use of classroom summative assessments is of interest. Teachers reported relying heavily on their own tests, quizzes, projects, etc., to make many informed decisions on practical applications. Shepherd (1989) says classroom assessments are probably more statistically unreliable but gather data about individual student learning in a much more accurate way than a standardized test. Darling-Hammond et al. (2005) and Shea et al. (2005) provide guidelines for what teachers should know and strategies on teaching teachers how to use summative assessments. This needs to become a focus in professional development and teacher preparation programs to ensure that teachers are creating these assessments in meaningful, useful manners that ensure the information gathered is accurate.

A final implication to consider is the extremely low use of formal assessments like school-level assessments, district-level assessments, and state level assessments. These are the standardized tests that administrators and policy makers tend to focus on (Shen & Cooley, 2008). If these are the assessments valued by the decision-makers, why are they not used in classroom practice? This is probably due to the fact that these exams are not directly connected to classroom content (Shen & Cooley, 2008), are inappropriate to gauge student learning (Marzano, 2003), are not meaningful to teachers' daily decisions (Creighton, 2001), and teachers do not feel they have the skills to interpret them meaningfully (Wayman & Stringfield, 2006). Greenberg and Walsh (2012) found that less than half of teachers view "outside data" like standardized tests as important. Even so, these assessments increasingly influence a teacher's performance reviews, salary, and evaluation (Linn, 2000). This paradox needs to be considered: Do we teach teachers what these assessments are meant for and how to use them appropriately? Do we teach them how to use formal assessments to influence their classroom practice and defend their practice via these assessments? Or do we consider it acceptable for teachers to continue to devalue formal assessments?

Limitations

There are limitations and potential concerns for this study. First of all, the measure is based on self-reported information. This is always a concern, as teachers may not have answered in a truthful manner or overestimated/underestimated their Knowledge and Use of assessments and data.

The sample size was smaller than desired. While it meets the requirement of 10 participants per item (DeVellis, 2003), more was hoped for when the samples were merged for ANOVA and correlation analyses. This was likely due to the timing of the surveys, which is another limitation. The calibration sample's data were collected in May at the end of North Dakota's school year, and while the sample's demographics were varied, this may have led to biased answers due to end-of-year fatigue. The snowball sample's data was collected during the summer months, which also may have led to more extreme scores from teachers willing to reflect on their practice outside of their standard school year. The other half of the validation sample's data were collected during August and September, during the beginning of the school year for North Dakota. While I believe this is the ideal time to collect this type of data, the first weeks of school can be intense and may have prevented some teachers from responding.

There was very limited racial diversity in this study. White females are the statistical majority demographic of teachers in the United States, so this is not unusual. This is further biased, as the majority of teachers from this study were from North Dakota, which is not a racially diverse state. The snowball sample and the pilot study had more diversity, but exploring differences across race still had to be considered as white or minority.

Finally, this study involved conducting a variety of statistical tests, in many occasions the same analysis was conducted multiple times. This can lead to a higher chance of making a type one error. Significance was set at 0.05 with no type one error inflation corrections included. For almost all findings, the p values associated were under 0.01 or even 0.001. Even so, using results from this study should be verified with new samples.

Future Research Recommendations

A scale that can capture teachers' classroom practice around data and assessment can be of value to future research. The tKUDA is recommended to be used to assess an entire school or school district in order to gauge classroom practices and then match these to the expectations based on policy or school or district goals. This would allow targeted professional development based on teacher feedback to meet teachers' needs and fill the gaps between policy and practice. It would be interesting to see if differences lie between professional learning groups, departments, or even whole schools. If differences are seen, a qualitative study could explore what the higher scoring groups are doing differently and if this is replicable for other situations.

Further study is needed to fully understand the tKUDA's dimensionality. This study supports a single, large factor with possible smaller facets, but recognizes that a larger sample may show multidimensionality.

Additionally, there is evidence from this study of differences in assessment use by content expertise and possibly even differences in the relationship between Knowledge and Use by content expertise. It is a logical assumption that an art teacher and a math teacher would use data in different ways and therefore may need different aspects of professional development specific to their needs. This hypothesis needs further exploration with quantitative support.

Finally, the relationship between Knowledge and Use was moderate at 0.47, with Knowledge explaining only 22% of Use. Future study is needed to understand other influences on data use. Increasing teachers' knowledge may not be the best or only way to increase use of data and assessments. Other important factors that may be influencing use and should be explored in future research are: teacher beliefs on data and assessment

(Coburn & Talbert, 2006; Slavit, Nelson, & Deuel, 2013; Tierney, 2006), self-efficacy when dealing with data and assessment (Skaalvik & Skaalvik, 2008), teacher resistance to change (Knight, 2009; Musanti & Pence, 2010; Zembylas, 2003), finding time in a teacher's day (Bartlett, 2004), and access/understanding of software to aid in assessment use (Shen & Cooley, 2008; Wayman & Stringfield, 2006; Wong, & Lam, 2007).

Conclusion: Final tKUDA Measure

In conclusion, support was found for the reliability and validity of the tKUDA that can be used to measure teachers' knowledge and use of the assessment process regarding data and assessments. The Knowledge factor shows strong support for reliability and unidimensionality. The Use factor shows adequate reliability, but person separation across items needs improvement. Dimensionality of the Use factor needs further exploration but evidence from Rasch analysis supports a single construct. The pilot study gave strong evidence of the tKUDA's effectiveness in identifying teachers' classroom practice regarding knowledge and use of the assessment process. The tKUDA can also capture practical applications of data use in teacher practice and what sources of data teachers rely on in order to execute those applications. When compared with policies, expectations, or goals, the tKUDA can identify teacher strengths and needs, which can then aid in targeted professional development opportunities.

The recommended final version of the tKUDA is provided in Appendix F. Item 2 ("communicating objectives using multiple methods") was problematic in multiple analyses for both Knowledge and Use factors and was therefore removed. Items 9 and 10 are about breaking down data to analyze it and were problematic in the Knowledge factor. Future studies could consider replacing these with different questions regarding how teachers analyze data. No other items are currently recommended for removal, but it is important to note that two other concerning items that should be reviewed in future studies: item 12 (“giving students feedback on how to improve”), and item 14 (“revising instruction immediately based on data”).

References

A common vision of great teaching: the Colorado Teacher Quality Standards (2013). In *Colorado Department of Education*. Retrieved July 20, 2014, from http://www.cde.state.co.us/sites/default/files/Colo%20Teacher%20Quality%20Standards %20Ref%20Guide%202.pdf

Arbuckle, J. L. (2006). Amos (Version 7.0) [Computer Program]. Chicago: SPSS.

Association for Childhood Education International Elementary Education Standards and Supporting Explanation (2007). In *Association for Childhood Education International*. Retrieved July 20, 2014, from http://acei.org/

Barnett, A. L. (2011). *Using data to inform instructional practice*. Distributed by ERIC Clearinghouse. (ED523328)

Bartlett, L. (2004). Expanding teacher work roles: a resource for retention or a recipe for overwork? *Journal of Education Policy*, *19*(5), 565-582.

Birrel, J., & Kee, R. (1996). Standardized testing and portfolio assessment: Rethinking the debate. *Reading, Research & Instruction, 35*(4), 285–298.

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

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences.* Mahwah, NJ: Erlbaum.

Brookhart, S.M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice.* *30*(1), 3–12.

Brookhart, S., Moss, C., & Long, B. (2008). *Formative assessment that empowers*.

Alexandria: Association for Supervision and Curriculum Development.

Bruster, B. G., & Peterson, B. R. (2013). Using critical incidents in teaching to promote reflective practice. *Reflective Practice, 14*(2), 170.

Cai, Y., & Lin, C. (2006). Theory and practice on teacher performance evaluation. *Frontiers of Education in China*, *1*(1), 29–39.

Cavalluzzo, L., Geraghty, T.M., Steele, J.L., Holian, L., Jenkins, F., Alexander, J.M., & Yamasaki, K.Y. (2014). Using Data to Inform Decisions: How Teachers Use Data to Inform Practice and Improve Student Performance in Mathematics: Results from a Randomized Experiment of Program Efficacy. CNA.

Chen, C., Crockett, M. D., Namikawa, T., Zilimu, J., & Lee, S. H. (2012). Eighth grade mathematics teachers' formative assessment practices in SES-different classrooms: A Taiwan study. *International Journal of Science and Mathematics Education, 10*(3), 553–579. doi:10.1007/s10763-011-9299-7

Coburn, C. E. & Talbert, J. E. (2006). Conceptions of Evidence Use in School Districts: Mapping the Terrain. American Journal of Education, 112, p. 469-495.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, N.J: L. Erlbaum Associates.

Creighton, T. B. (2001). Data analysis in administrators' hands: An oxymoron? *The School Administrator*, *58*(4), 6–11.

Creswell, J. W. (2007). *Qualitative inquiry and research design; Choosing among five approaches* (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Creswell, J. W., & Plano-Clark, V. L. (2011). Designing and conducting mixed methods research (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Darling-Hammond, L., Bransford, J., & National Academy of Education. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco, CA: Jossey-Bass.

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

Determining final effectiveness ratings using the Colorado state model evaluation system (2013). In *Colorado Department of Education*. Retrieved from http://www.cde.state.co.us/sites/default/files/Determining%20Final%20Rating%2008-01-2013.pdf

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Thousand Oaks, CA: SAGE Publications.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston, MA: DC Heath.

Earl, L. M., & Katz, S. (2006). *Leading schools in a data-rich world: Harnessing data for school improvement*. Thousand Oaks, CA: Corwin Press.

Frohbieter, G., Greenwald, E., Stecher, B., & Schwartz, H. (2011). *Knowing and doing: What teachers learn from formative assessment and how they use the information.* (CRESST Report 802). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Froschauer, L. (2013). Seamless assessment. *Science and Children, 51*(3), 6.

Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology*. Englewood Cliffs, N.J: Prentice-Hall.

Good, M. E. (2007). *Differentiated instruction: Principles and techniques for the elementary grades.* Distributed by ERIC Clearinghouse, 1–34.

Greenberg, J., & Walsh, K. (2012). What teacher preparation programs teach about K–12 assessment: A review of coursework on K–12 assessment from a sample of teacher preparation programs. Distributed by National Council on Teacher Quality. Accessed online at http://www.nctq.org/dmsView/What_Teacher_Prep_Programs_Teach_K-12_Assessment_NCTQ_Report

Hailaya, W., Alagumalai, S., & Ben, F. (2014). Examining the utility of Assessment Literacy Inventory and its portability to education systems in the Asia Pacific region. *Australian Journal of Education*, *58*, 297–317.

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). Using student achievement data to support instructional decision making. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Heritage, M., & Bailey, A.L. (2006). Assessing to teach: An introduction. *Educational Assessment. 11*(34), 145–148. DOI:10.1080/10627197.2006.9652987

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan. 89*(2).

Herman, J. L. (1992). What research tells us about good assessment. *Educational Leadership, 49*(8), 74–78.

Hoover, N. R., & Abrams, L. M. (2013). Teachers' instructional use of summative student assessment data. *Applied Measurement in Education, 26*(3), 219–231.

Hunt, E., & Pellegrino, J. W. (2002). Issues, examples, and challenges in formative assessment. *New Directions for Teaching and Learning, 89*, 73–85. doi:10.1002/tl.48

Jennings, J. L. (2012). The effects of accountability system design on teachers' use of test score data. *Teachers College Record, 114*, 1–23.

Kellaghan, T., Stufflebeam, D. L., & Credo Reference Firm (2003). *International handbook of educational evaluation*. Dordrecht [Netherlands]: Kluwer Academic Publishers.

Kline, A. J. (2013). *Effects of formative assessment on middle school student achievement in mathematics and reading* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 1538121).

Knight, J. (2009). What can we do about teacher resistance. *Phi Delta Kappan,90*(7), 508-513.

Linacre, J. M. (2015). *A user's guide to Winsteps/Ministep: Rasch-model computer programs*. Chicago: Winsteps.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4–16. doi:10.2307/1177052

Little, J. W. (2003). Inside teacher community: Representations of classroom practice.

Teachers College Record, 105(6), 913–945.

Little, J. W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education, 118*, 143–166.

Lysaght, Z., & O’Leary, M. (2013). An instrument to audit teachers’ use of assessment for learning. *Irish Educational Studies, 32*, 217–232.

Madaus, G. F., & O’Dwyer, L. M. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappan, 80*(9), 688–698.

Marsh, J. A., McCombs, J., & Martorell, F. (2009). How instructional coaches support data-driven decision making: Policy implementation and effects in Florida middle schools. *Educational Policy, 24*(6), 872–907.

Marzano, R. (2003). Using data: Two wrongs and a right. *Educational Leadership, 60*(5), 56–60.

McDonald, M. (2002). The perceived role of diploma examinations in Alberta, Canada. *Educational Research, 96*(1), 21–36.

MacFarlane, B. (2012). Differentiating teacher professional development with design. *Understanding Our Gifted, 24*(2), 9–14.

McLeod, S. (2009). Data-driven teachers. In *Innovative Teacher Thought Leaders.* Retrieved from https://datacenter.spps.org/uploads/ThoughtLeaders_DDDM_May05.pdf/

McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation, 7*(8), 1–9.

Means, B., Gallagher, L., Padilla, C., & U.S. Department of Education, Office of

Planning, Evaluation, and Policy Development (2010). Use of education data at the local level from accountability to instructional improvement. Washington, D.C.

Mendenhill, W., & Sincich, T. (2012). *A second course in statistics: regression analysis*. Prentice Hall: Boston.

Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge & application of classroom assessment concepts: Development of the "Assessment Literacy Inventory."* Paper presented at the annual meeting of the American Educational Research Association in Montreal.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education, 79*(4), 33–53. doi:10.1207/s15327930pje7904_3

Monpas-Huber, J. B. (2010). Explaining teachers' instructional use of state assessment data: a multilevel study of high school teachers in Washington state. *Journal of School Leadership, 20*, 208–237.

Musanti, S. I., & Pence, L. (. (2010). Collaboration and teacher development: Unpacking resistance, constructing knowledge, and navigating identities. *Teacher Education Quarterly, 37*(1), 73-89.

National Council for Social Studies Teachers (2002). Retrieved from http://www.socialstudies.org/standards/teacherstandards

National Education Association of the United States (1983). Teachers' views about student assessment. *Division of Instruction and Professional Development.* Washington, D.C.: The Association.

National Research Council (1999). *Uncommon measures: Equivalence and linkage among educational tests*. Washington, D.C.: Author.

National Research Council (2001). *Knowing what students Know: The science and design of educational assessment.* Committee on the Foundations of Assessment. Pelligrino, J., Chudowsky, N., & Glaser, R., editors. Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, *22*(2), 155–175.

NCATE/NCTM Program Standards (2003). In *National Council of Teachers of Mathematics*. Retrieved from http://www.ncate.org/ProgramStandards/NCTM/NCTMSECONStandards.pdf

NCTE/NCATE standards for Initial Preparation of Teachers of Secondary English Language Arts Grades 7–12 (2012). Retrieved from http://www.ncte.org/library/nctefiles/groups/cee/ncate/approvedstandards_111212.pdf

Nelson, T. H., Slavit, D., & Deuel, A. (2012). Two dimensions of an inquiry stance toward student-learning data. *Teachers College Record, 114*(8), 1–4.

NSTA Preservice Science Standards (2012). Retrieved from http://www.nsta.org/preservice/docs/2012NSTAPreserviceScienceStandards.pdf

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257. doi:10.3102/01623737026003237

Perry, M. L. (2013). *Teacher and principal assessment literacy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3568118).

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment Competencies of Teachers: A National Survey. *Educational Measurement Issues and Practice, 12*(4), 10–12. doi:10.1111/j.1745-3992.1993.tb00548.x

Praxis exams (2014). Retrieved from https://www.ets.org/praxis/co

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago.

Rubric for evaluating Colorado teachers (2013). In *Colorado Department of Education*. Retrieved from http://www.cde.state.co.us/sites/default/files/TeacherRubric.pdf

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education, 26*(3), 482–496.

Shea, M., Murray, R., & Harlin, R. (2005). *Drowning in data?: How to collect, organize, and document student performance*. Portsmouth, NH: Heinemann.

Shen, J., & Cooley, V. E. (2008). Critical issues in using data for decision-making. *International Journal of Leadership in Education, 11*(3), 319–329.

Shepard, L., Davidson, K., & Bowman, R. (2011). *How middle-school mathematics* *teachers use interim and benchmark assessment data.* (CRESST Report 807). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Shepard, L. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice, 28*(3), 32–37.

Shepard, L. (1989). Why we need better assessments. *Educational Leadership, 46*(7), 4–9.

Shepard, L. (1979). Purposes of assessment. *Studies in Educational Evaluation. 5*, 13–26.

Siegel, M. A. & Wissehr C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, *22*, 371–391.

Sinkovics, R. R., Penz, E., & Ghauri, P. N. (2008). Enhancing the trustworthiness of qualitative research in international business. *Management International Review. 48*(6), 89–714. DOI: 10.1007/s11575-008-0103-z

Skaalvik, E. M., & Skaalvik, S. (2008). Teacher self-efficacy. *Self-processes, Learning, and Enabling Human Potential: Dynamic New Approaches*, 223.

Slavit, D., Nelson, T. H., & Deuel, A. (2013). Teacher groups' conceptions and uses of student-learning data. Journal of Teacher Education, 64(1), 8-21

Standards for Science Teacher Preparation (2003). In *National Science Teachers Association*. Retrieved from www.nsta.org/preservice/

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: review of research. *Psychology in the Schools, 42*(8).

Stiggens, R., & DuFour, R. (2009). Maximizing the power of formative assessments. *Phi Delta Kappan, 90*(9), 640.

Stiggens, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*, 758–69.

Stiggins, R. J. (1991). Assessment literacy. *The Phi Delta Kappan*, *72*(7), 534–539.

Stronge, J. H., & Grant, L. W. (2009). *Student achievement goal setting: Using data to improve teaching and learning*. Larchmont, NY: Eye On Education.

Supovitz, J. (2012). Getting at student understanding; The key to teachers' use of test data. *Teachers College Record*, *114*, 1.

Taylor, E. S., & Tyler, J. H. (2012). Can teacher evaluation improve teaching? *Education Next*, *12*(4), 78–84.

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics, 29*, 21–36. doi:10.1017/S0267190509090035

Tomlinson, C. A. (2000). *Differentiation of instruction in the elementary grades.* Distributed by ERIC Clearinghouse.

Tok, S., & Dolapçioglu, S. D. (2013). Reflective teaching practices in Turkish primary school teachers. *Teacher Development, 17*(2), 265.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal, 27*(5), 615-631.

Wayman, J. C., & Stringfield, S. (2006). Data use for school improvement: School Practices and research perspectives. *American Journal of Education, 112*(4), 463-468.

Walsh, K. (2003). *After the test: How schools are using data to close the achievement gap.* Bay Area School Reform Collaborative, San Francisco.

Walvoord, B. E. F., & Anderson, V. J. (2009). *Effective grading: A tool for learning and assessment in college*. Hoboken: John Wiley & Sons, Inc.

Walvoord, B. E. F., & Banta, T. W. (2010). *Assessment clear and simple: A practical guide for institutions, departments, and general education*. Hoboken: John Wiley & Sons, Inc.

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Williams, D. (2011). What is assessment for learning? *Studies in Educational Evaluation. 37*, 3–14.

Williams, D. (2013). *What formative assessment is (and isn't) and practical techniques for implementing formative assessments* [Online Lecture]. Retrieved from http://info.nwea.org/FY2012WinterCampaignKLTWebinar2v2_On-demandThankYou.html

Wilson, R. B. (2009). *The effect of a continuous quality formative assessment program on middle school student mathematics achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No 3378822).

Wong, O. K., & Lam, M. (2007). *Using data analysis to improve student learning:*

Toward 100% proficiency. Lanham, MD: Rowman & Littlefield Education.

Wright, B. D. (1994). Comparing factor analysis and Rasch measurement. *Rasch Measurement Transactions*, *8*(1), 350-357.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8,* 370.

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, A., Brandon, P. R., Furtak, E. M., Tomita, M. K., & Young, D. B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, *21*(4), 335–359.

Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education*, *112*(4), 521–548.

Zembylas, M. (2003). Interrogating "Teacher Identity": Emotion, resistance, and self-formation. *Educational theory*, *53*(1), 107-127.

Appendix A

Interview Questions

Focus Group

- What should a teacher know about data and assessments? Please create a list all of the things a teacher should know about data and assessments.
- What should a teacher be able to do with data and assessments? Please list all of the things a teacher should be able to do with data and assessments.
- How do these connect to each other? Discuss and brainstorm connections together. Show and compare the matching based on standards and literature. What is missing?
- How do your own beliefs affect how you use data?
- Is data used to Prove or Improve learning? Do you think all data are used to Prove/Improve learning, or do these ideas change when considering type of data or assessment?
- *Show and explain the conceptual framework.* Open discussion: What are your thoughts? Does this hold true to you? Do the arrows make sense or should they be modified?
- Discuss survey idea (Appendix B). Do these ideas seem true? What should I change? What should I add?
- What terms need to be defined? What terms need examples to help explain them?
- Thinking back to the whole discussion. Is there anything that didn't feel right? What is missing? What needs to be expanded?

Expert Interviews

- Discussion of conceptual framework. What are your thoughts? Does this hold true to you? Do the arrows make sense or should they be modified? How do you think Knowledge is related to Belief? Do you think all data are used to Prove/Improve learning, or do these ideas change when considering the type of data or assessment?
- Discussion of survey structure. What are your thoughts and opinions of the factor order (Belief, then Knowledge, then Use)? On the item structure and item wording?

Cognitive Interviews

- Go through each item and for some items discuss: What does this mean to you? How are you thinking about this? How might you answer? Why did you select the response you did?
- What terms need to be defined? What terms need examples to help explain them?

- Is there anything that didn’t feel right? What is missing? What needs to be expanded?
- Was there anything that seemed redundant or should be deleted?

Appendix B

tKUDA Survey Items

Part 1: tKUDA

Knowledge of the Assessment Process

This section of the survey is attempting to measure teachers' knowledge of conducting the key parts of the assessment process. The assessment process includes setting learning objectives, choosing a variety of assessment strategies, analyzing information on student learning, reflecting and revising instructions, and of course communicating this with students throughout the process.

In this section, please answer each in regards to what you *know*, regardless of if you are able to put it into practice. Rate each item based on how strongly you agree with each statement using:

Agreement Scale (1-Strongly Disagree, 2-Disagree, 3-Agree, 4-Strongly Agree)

1. I know how to set specific learning goals/objectives.
2. I know how to communicate my learning goals to students **using multiple methods** (like posting in the room, via PowerPoint, verbally, having students write it, having students discuss it, reviewing it, highlighting it during lesson, checking for understanding specifically based on objective, etc.).
3. I know how to pre-assess my students for prior knowledge.
4. I know how to choose or create an assessment strategy that will measure my specific learning objective (like a task, project, discussion, exit slip, quiz, test, etc.).
5. I know how to use a variety of assessment techniques.
6. I know how to effectively use assessments to show students' *thinking*, not just their answer.
7. I know how to evaluate evidence from assessments in order to **prove** student learning.
8. I know how to evaluate evidence from assessments in order to **improve** student learning.
9. I know how to read data (typically from a standardized exam) when it is broken down for me based on gender, race/ethnicity, IEP, GT, content standard, etc.
10. I know how to break down **my own** assessments based on gender, race/ethnicity, IEP, GT, content standard, etc.
11. I know how to give students specific feedback on **what** they need to improve.
12. I know how to give students specific feedback on **how** to improve.
13. I know how to reflect on my instructional practices based on evidence from my assessment techniques (observations, questioning, quizzes, tests, projects, etc.).

14. I know how to revise my instructional practices **immediately (on the fly)** based on evidence from assessments (observations, questioning, quizzes, tests, projects, etc.).
15. I know how to revise my instructional practices **for the next year** based on evidence from assessments (observations, questioning, quizzes, tests, projects, etc.).

Use of the Assessment Process

This section of the survey is attempting to measure how frequently teachers are able to use different aspects of the assessment process. The assessment process includes setting learning objectives, choosing a variety of assessment strategies, analyzing information on student learning, reflecting and revising instructions, and of course communicating this with students throughout the process.

In this section, please answer each in regards to how often you are able to do each item. Some items may be considered differently based on multiple classes you teach, but please answer based on whatever you consider a *typical* day or situation. Rate each item by finishing the sentence using:

0-Never, 1-Yearly, 2-Quarterly, 3-Monthly, 4-Weekly, 5-Daily

1. I set specific learning goals/objectives…..
2. I *typically* communicate my learning goals to students **using multiple methods** (posting in the room, via PowerPoint, verbally, having students write it, having students discuss it, reviewing it, highlighting it during lesson, checking for understanding specifically based on objective, etc.)….
3. I pre-assess my students….
4. I choose or create an assessment strategy to measure a specific learning goal (like a task, project, discussion, quiz, test, etc.) ….
5. I use a *variety* of assessment techniques….
6. I effectively use assessments to show students' *thinking*, not just their answer…
7. I evaluate evidence from assessments in order to **prove** student learning….
8. I evaluate evidence from assessments in order to **improve** student learning….
9. I *typically* read data that is broken down for me based on gender, race/ethnicity, IEP, GT, content standard, etc….
10. I break down my own assessments based on gender, race/ethnicity, IEP, GT, content standard, etc….
11. I give students specific feedback on **what** they need to improve….
12. I give students specific feedback on **how** to improve…
13. I reflect on my instructional practices **based on** evidence from assessment techniques (observations, questioning, quizzes, tests, projects, etc.)…..
14. I revise my instructional practices **immediately (on the fly)** based on evidence from assessment techniques (observations, questioning, quizzes, tests, projects, etc.)…..

15. I revise my instructional practices **for the next year** based evidence from assessment techniques (observations, questioning, quizzes, tests, projects, etc.)…..

Practical Application of Different Types of Assessments

This section of the survey tries to understand how often teachers use the various types of assessment strategies and what they use them for in their teaching practice. One of the challenges of understanding how teachers use the assessment process is due to the many types of assessment strategies and numerous practical applications of data gained from these strategies. Some strategies are more formal, documented assessments (like a state exam or a unit test), while others are more informal techniques (like observations or exit slips). There are multiple practical applications to use student learning data and the way a teacher uses the information from these assessments can depend on the type of assessment data being considered.

In this section, please answer each in regards to things you *typically* do based on each assessment type. Please complete each sentence by marking each type of assessments you use for that application. For example, I might use state assessments for one application, but planned formative assessments for another application. There is no expectation that each practical application is appropriate for each type of assessment, especially when considering various content or grade levels.

Options to check multiple choice (across columns):

- My **planned formative** assessments *(exit slips, fingers 1-5, white boards, discussions, observations, etc.)*
- My "on the fly," **in the moment formative** assessments *(examples column 2)*
- My **classroom summative** assessments *(test, project quiz, rubric, performance)*
- **School-Level** assessments *(common assessments, finals, etc.)*
- **District-Level** assessments *(benchmark or interim tests)*
- **State- level** assessments *(TCAP, CMAS)*
- I **never** do this or do not use data for this

Items (down rows):

1. I inform (drive) my instruction using…
2. I set new learning goals using…
3. I differentiate instruction using…
4. I create student groups using…
5. I reteach or review content using…
6. I identify gaps in learning or target skills using…
7. I reflect on and revise instruction based on data (modify and adjust plans) using…
8. I gauge my students' engagement level using…
9. I get a "feeling for" incoming students using…
10. I give **parents** feedback of student learning using…
11. I give **students** feedback on their learning using…

12. I facilitate student goal setting using…
13. I identify student strengths and weaknesses using…
14. I explore patterns across students using…
15. I demonstrate/prove student achievement using…
16. I track student growth and/or conduct progress monitoring using…
17. I monitor target students using…
18. I analyze information by specific item or content standard using…
19. I analyze information by student ethnicity, gender, proficiency levels, IEP, GT, etc. using…
20. I decide or recommend student placement in programming or specific classes using…
21. I predict students' future scores using…
22. I triangulate learning (using this as one of many sources to show student learning) using…

Part 2: Demographic Information

1. Race/ethnicity (White, Hispanic/Latino, African-American, Asian, American Indian/Alaska Native, Native Hawaiian/ Pacific Islander, Other, Prefer not to answer)
2. Gender (Male, Female, Other, Prefer not to answer)
3. Years teaching (1–50)
4. Content expertise, check all that apply (Math, Science, Language Arts, Social Studies, Music, Art, Physical Education, Computer/Technology, Foreign Language, Other)
5. What grade level(s) do you teach? (K–12)

Part 3: "Using Data to Inform Decisions" survey

In school year 2014-2015, how often have you use data for each of the following purposes?

1 - Never, 2 - A few times, 3 - Once or twice a month, 1 - Once a week or more

In 2012/13, I used data...
1. to inform curriculum changes
2. to identify individual skill gaps for individual students
3. to determine whether your class or individual students were ready to move on to the next instructional unit
4. to evaluate promising classroom practices
5. to decide to give your students test-taking practice
6. to estimate whether your students would make adequate yearly progress (AYP)
7. to track standardized test scores by grade
8. to track individual student test scores
9. to track other measures of student progress

10. to inform student placement in courses or special programs

Appendix C

Informed Consent

Survey Informed Consent Form

Note: This will appear in the email sent to participants and will be the first section of the Qualtix survey. Participants who do not agree will be exited from the survey.

You are invited to participate in a research study that is working to create a measure of teachers' knowledge and use of data and assessment. Your participation is completely voluntary, but it is very important as it will help capture specific teacher-perspectives and teacher-voices.

If you agree to be part of the research study, you will be asked to complete an online survey about your experiences as a K–12 teacher. This survey will take 20 to 40 minutes to complete. The survey questions will ask about your knowledge and use of data and assessment.

While you may not receive any direct benefit for participating, I hope that this project will contribute to decisions being made by school districts, teacher preparation programs, and policy makers to improve requirements and professional development supports by giving teachers more of a voice.

As the researcher, I will not be able to link your survey response to you. The survey software keeps your identifying information separate from the answers you provide to the survey; it is completely anonymous. I do plan to publish the results of this study, but will not include any information that would identify you. Individual results will be kept anonymous from your administration and school district and will not impact your employment.

Participation is completely voluntary and you may stop at any time. You may choose to not answer a specific question or section of the survey by clicking NEXT without providing an answer.

If you have any questions at all about this study, please feel free to contact Courtney Vidacovich Tobiassen. If you have any concerns or complaints about how you were treated during the research sessions please contact Dr. Susan Sadler, Chair, Institutional Review Board for Human Subjects, University of Denver, 303-871-3454, or du-irb@du.edu, or write to the University of Denver, Office of Sponsored Programs, 2199 S. University Blvd., Denver, CO 80208-2121. Thank you.

"I have read and understand the above description of the study that is working to create a measure of teachers' knowledge and use of data and assessment. I have asked for and received a satisfactory explanation for any language I did not fully understand. I have had the chance to ask any questions I have about my participation. I agree to participate in the study, and I understand that I can withdraw my consent at any time. I have received a copy of this consent form."

If you agree to the above statement please click YES and continue to the survey. If not, click NO and it will exit you from the survey. Thank you for your time and aid in this project.

Click here to go to the survey: ___________________

Appendix D

Expert Review Response Form

Construct definitions

Knowledge -

Use -

Potential Items	**<u>Clarity Rating</u>** (1-Not Clear to 4-Very Clear)	**<u>Representativeness to Domain</u>** (1-Not appropriate to 4-Very appropriate for this domain)	**<u>Item Difficulty</u>** (4- Difficult to 1-Easy)

Please comment on the following:

1. What should be defined and/or needs examples?
2. Scale appropriateness for each domain:
3. Comprehensiveness of items:
4. Overall construct and ordering:
5. Suggestions of additional items, item re-wordings, or item deletions:
6. Any other thoughts or concerns:

Appendix E

Supporting Tables

Table 1

tKUDA All Items Eigenvalues

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	10.08	33.60	33.60
2	3.47	11.55	45.15
3	1.82	6.07	51.22
4	1.41	4.68	55.90
5	1.32	4.41	60.31
6	1.21	4.05	64.35

Table 2

tKUDA All Items Two Factor Rotated Solution

	Component	
Item	1	2
1. I know how to set specific learning goals/objectives.	.72	.16
2. I know how to communicate my learning goals to students using multiple methods	.56	.16
3. I know how to assess my students for prior knowledge.	.78	.04
4. I know how to choose or create an assessment strategy that will measure my specific learning objective	.77	.05
5. I know how to use a variety of assessment techniques.	.75	-.01
6. I know how to effectively use assessments to show students' thinking, not just their answer.	.71	.09
7. I know how to evaluate evidence from assessments in order to prove student learning.	.74	.10
8. I know how to evaluate evidence from assessments in order to improve student learning.	.75	.15
9. I know how to read data (typically from a standardized exam) when it is broken down for me based on gender, race/ethnicity, IEP, GT, content standard, etc.	.57	.24

10. I know how to break down results from my own assessments based on gender, race/ethnicity, IEP, GT, content standard, etc.	.66	.19
11. I know how to give students specific feedback on what they need to improve.	.76	.18
12. I know how to give students specific feedback on how to improve.	.71	.26
13. I know how to reflect on my instructional practices based on evidence from my assessment techniques	.75	.22
14. I know how to revise my instructional practices immediately (on the fly) based on evidence from assessments	.70	.33
15. I know how to revise my instructional practices for the next year based on evidence from assessments	.75	.28
1. I set specific learning goals/objectives	.01	.66
2. I typically communicate my learning goals to students using multiple methods	.03	.63
3. I assess my students for prior knowledge	.02	.49
4. I choose or create an assessment strategy to measure a specific learning goal	.13	.61
5. I use a variety of assessment techniques	.16	.54
6. I effectively use assessments to show students'; thinking, not just their answer	.25	.56
7. I evaluate evidence from assessments in order to prove student learning	.16	.64
8. I evaluate evidence from assessments in order to improve student learning	.28	.62
9. I typically read data that is broken down for me based on gender, race/ethnicity, IEP, GT, content standard, etc.	.35	.43
10. I break down results from my own assessments based on gender, race/ethnicity, IEP, GT, content standard, etc.	.32	.47
11. I give students specific feedback on what they need to improve	-.03	.59
12. I give students specific feedback on how to improve	.14	.62
13. I reflect on my instructional practices based on evidence from assessment techniques	.28	.61
14. I revise my instructional practices immediately (on the fly) based on evidence from assessment techniques	.08	.50
15. I revise my instructional practices for the next year based evidence from assessment techniques	.18	.45

Table 3

Knowledge Factor Eigenvalues

	Eigenvalues	% of Variance	Cumulative %
1	5.38	35.87	35.87
2	1.54	10.28	46.15
3	1.25	8.30	54.45
4	1.16	7.75	62.20

Table 4

Knowledge Factor Item Loadings Unrotated Solution

	Component			
	1	2	3	4
Setting learning objectives	.66	-.26	-.32	.06
Communicating objectives	.62	-.27	-.28	.08
Assessing prior knowledge	.50	-.03	-.09	-.16
Choosing assessment from objective	.64	-.07	-.34	-.23
Using a variety of assessments	.58	-.16	-.16	-.19
Showing student thinking	.64	.14	-.36	-.07
Using assessments to prove learning	.69	.13	-.12	-.20
Using assessments to improve learning	.70	.11	-.03	-.06
Reading disaggregated data	.54	.66	.21	-.11
Disaggregating their own data	.53	.67	.28	-.11
Giving feedback on what to improve	.56	-.44	.54	-.27
Giving feedback on how to improve	.64	-.35	.56	-.18
Reflecting using data	.67	-.07	.08	.46
Revising instruction immediately	.47	-.20	.19	.61
Revising instruction later	.48	.28	.03	.53

Table 5

Use Factor Eigenvalues

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	5.38	35.87	35.87
2	1.54	10.28	46.15
3	1.25	8.30	54.45
4	1.16	7.75	62.20

Table 6

Use Factor Item Loadings Unrotated Solution

	Component			
	1	2	3	4
Setting learning objectives	.66	-.26	-.32	.06
Communicating objectives	.62	-.28	-.28	.08
Assessing prior knowledge	.49	-.03	-.09	-.16
Choosing assessment from objective	.63	-.07	-.34	-.23
Using a variety of assessments	.58	-.16	-.16	-.19
Showing student thinking	.63	.14	-.36	-.07
Using assessments to prove learning	.69	.13	-.12	-.20
Using assessments to improve learning	.69	.11	-.03	-.06
Reading disaggregated data	.53	.66	.21	-.11
Disaggregating their own data	.53	.67	.28	-.11
Giving feedback on what to improve	.56	-.44	.54	-.27

Giving feedback on how to improve	.64	-.35	.56	-.18
Reflecting using data	.67	-.07	.08	.46
Revising instruction immediately	.47	-.20	.19	.61
Revising instruction later	.48	.28	.03	.53

Appendix F

Teachers Knowledge of Data and Assessment (tKUDA)

tKUDA: Knowledge of the Assessment Process Items

Scale of 1-Disagree, 2-Somewhat Agree, 3-Agree, 4-Strongly Agree

1. I know how to set specific learning goals/objectives.
2. I know how to pre-assess my students for prior knowledge.
3. I know how to choose or create an assessment strategy that will measure my specific learning objective (like a task, project, discussion, exit slip, quiz, test, etc.).
4. I know how to use a variety of assessment techniques.
5. I know how to effectively use assessments to show students' *thinking*, not just their answers.
6. I know how to evaluate evidence from assessments in order to prove student learning.

7. I know how to evaluate evidence from assessments in order to improve student learning.
8. I know how to read data when it is broken down for me based on gender, race/ethnicity, IEP, GT, content standard, etc.
9. I know how to break down my own assessments based on gender, race/ethnicity, IEP, GT, content standard, etc.
10. I know how to give students specific feedback on *what* they need to improve.
11. I know how to give students specific feedback on *how* to improve.
12. I know how to reflect on my instructional practices based on evidence from my assessment techniques (observations, questioning, quizzes, tests, projects, etc.).
13. I know how to revise my instructional practices immediately (on the fly) based on evidence from assessments (observations, questioning, quizzes, tests, projects, etc.).
14. I know how to revise my instructional practices for the next year based on evidence from assessments (observations, questioning, quizzes, tests, projects, etc.).

tKUDA: Use of the Assessment Process Items

Scale: 1-Never to Yearly, 2-Quarterly to Monthly, 3-Weekly, 4-Daily

1. I set specific learning goals/objectives…
2. I pre-assess my students…

3. I choose or create an assessment strategy to measure a specific learning goal (like a task, project, discussion, quiz, test, etc.)…
4. I use a *variety* of assessment techniques…
5. I effectively use assessments to show students' *thinking*, not just their answers…
6. I evaluate evidence from assessments in order to prove student learning…
7. I evaluate evidence from assessments in order to improve student learning…
8. I *typically* read data that is broken down for me based on gender, race/ethnicity, IEP, GT, content standard, etc.…
9. I break down my own assessments based on gender, race/ethnicity, IEP, GT, content standard, etc.…
10. I give students specific feedback on *what* they need to improve…
11. I give students specific feedback on *how* to improve…
12. I reflect on my instructional practices based on evidence from assessment techniques (observations, questioning, quizzes, tests, projects, etc.)…
13. I revise my instructional practices immediately (on the fly) based on evidence from assessment techniques (observations, questioning, quizzes, tests, projects, etc.)…
14. I revise my instructional practices for the next year based evidence from assessment techniques (observations, questioning, quizzes, tests, projects, etc.)…

tKUDA: Practical Application of Different Types of Assessments

Options to check multiple choice (across columns):

- My planned formative assessments *(exit slips, fingers 1–5, white boards, discussions, observations, etc.)*
- My "on the fly," in the moment formative assessments *(examples column 2)*
- My classroom summative assessments *(tests, project quizes, rubrics, performances)*
- School-Level assessments *(common assessments, finals, etc.)*
- District-Level assessments *(benchmark or interim tests)*
- State-Level assessments
- I never do this or do not use data for this

Items (down rows):

1. I inform (drive) my instruction using…
2. I set new learning goals using…
3. I differentiate instruction using…
4. I create student groups using…
5. I reteach or review content using…
6. I identify gaps in learning or target skills using…
7. I reflect on and revise instruction based on data (modify and adjust plans) using…
8. I gauge my students' engagement level using…
9. I get a "feeling for" incoming students using…
10. I give parents feedback of student learning using…
11. I give students feedback on their learning using…
12. I facilitate student goal setting using…

13. I identify student strengths and weaknesses using…
14. I explore patterns across students using…
15. I demonstrate/prove student achievement using…
16. I track student growth and/or conduct progress monitoring using…
17. I monitor target students using…
18. I analyze information by specific item or content standard using…
19. I analyze information by student ethnicity, gender, proficiency levels, IEP, GT, etc., using…
20. I decide or recommend student placement in programming or specific classes using…
21. I predict students' future scores using…
22. I triangulate learning (using this as one of many sources to show student learning) using…