5-2016

Examination of the NU Data Knowledge Scale

Jonathon David Sikorski
University of Nebraska-Lincoln, jonathon.sikorski@gmail.com

EXAMINATION OF THE NU DATA KNOWLEDGE SCALE

- by-

Jonathon Sikorski

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Psychological Studies in Education

(School Psychology)

Under the Supervision of Professor Beth Doll

Lincoln, Nebraska

May, 2016

EXAMINATION OF THE NU DATA KNOWLEDGE SCALE

Jonathon Sikorski, Ph.D.

University of Nebraska, 2016

Advisor: Beth Doll

There is a pervasive need for school systems to empirically and reliably assess the data literacy and data use skills of their educators (Ingram et al., 2004). With the federal government holding states and school districts accountable for student achievement and the increased emphasis of high stakes testing in schools, it is also critical to be able to precisely and accurately assess the skill and knowledge level of educators by measuring their overall data literacy. Disentangling these skills and abilities is a difficult task and, to date, no empirical measure of data literacy has been established. A strong measure of data literacy would have an empirical evidence base, have items that are reliable and internally consistent, and be recognized by experts in the scientific community as being empirically valid and conceptually sound.

The focus of this dissertation is the development of the NU Data Knowledge Scale: A measure of teachers' data use skills and knowledge. The psychometric properties of the NU Data Knowledge Scale were thoroughly examined in this dissertation. First, the test items were based off the databasics, and were independently categorized by subject matter experts. The measure was revised based off of the recommendation of the subject matter experts. The survey was sent to 215 rural Nebraskan teachers along with a demographics section and "Comfort with Data Use" questionnaire. The psychometric properties of the measure were discussed that related the internal consistency, item-total correlations, item difficulty, and item discrimination. The

dimensionality of the scale was explored using weighted least means squares analysis and the factor solution was determined by computing a parallel analysis. Fourteen predictors of teacher data literacy were then analyzed through an all possible regression procedure and the top model was chosen based off the Mallow's Cp and adjusted R^2.

Overall, the NU Data Knowledge Scale was found to be a single factor measure of data literacy. The predictors included in this model, though significant, did not provide practical significance in predicting scores on the measure. The limitations of the study, direction for future search, and implications for future practice are discussed.

TABLE OF CONTENTS

Figures

Appendix

CHAPTER ONE: IMPORTANCE

The demands on educators to look beyond student tests scores have never been higher. Educators today must be able to screen, give benchmark assessments, analyze graphs, and integrate data from multiple sources to describe the ability level of students (Jacobs, Gregory, Hoppey, & Yendol-Hoppey, 2009). Educational reform stemming from the No Child Left Behind Act (NCLB; 2001) has pushed educators away from using subjective measures of student progress (i.e., observations, gut feelings, and opinions) towards more rigorous, empirical approaches (Dunn, Airola, Lo, & Garrison, 2013; Stockard, 2013). NCLB legislation mandates students achieve specific benchmarks on high stakes assessments which increases pressure on schools to use student data well (Thorndike & Thorndike-Christ, 2010; Wayman, Spikes, & Volonnino, 2013). As a result, school districts seek out teacher development programs and assessment practices that require teachers to examine, gather, and interpret student data effectively (Mandinach & Gummer, 2013b; Qin & D'Ignazio, 2010). Indeed, educators who are adequately trained in effective data use strategies are better able to make educational decisions for students that are empirically based (Wayman et al., 2013).

In educational settings, data literacy research has become a focal point for discussion as school systems attempt to become more data driven and evidence based (Duncan, 2009). High stakes testing has put pressure on educators to meet performance-based standards at the district, state, and federal level (Mertler & Campbell, 2005). Unfortunately, data use strategies and skills have not translated into educator training and preparation programs (Dunn et al., 2013). Consequently, school districts have been

responsible for training a majority of their teaching staff in the latest evidence-based approaches (Wayman et al., 2013).

Data is an umbrella term used to describe a variety of information gathering tools. In school settings, student data is multifaceted and can include: archival student data such as student records, Individualized Education Program (IEP) goals, and standardized test scores; permanent products such as classroom assignments or projects; observational data such as student behavior; and cognitive ability scores from test scores, intelligence tests, and standardized assessments (Choppin, 2002). Teachers who are well versed in collecting and organizing student data from multiple sources are better able to differentiate instruction to meet the needs of their students and schools (Means, Chen, DeBarger, & Padilla, 2011). One promising teacher training program that taught teachers how to use multiple sources of data to strengthen student success was the NU Data Intervention (Doll et al., 2010).

The NU Data Intervention

The NU Data Intervention is a professional development program with the purpose of strengthening teachers' use of student data (Doll, Horn, & Shope, 2010; Sikorski, Doll, Thomas, Franta, & Kenney, 2013). The NU Data intervention is salient to data literacy research as it provides teachers opportunities to implement the six DataBasics through developmental instruction via weekly check-ins, site visits, and three seminars (Doll, Franta, Thomas, Chapla, & Sikorski, 2014; Sikorski, Franta, & Doll, 2014). This blended approach helped the researchers overcome the distance barriers of consulting with rural schools. As a result, teachers were able to develop data literacy and matriculate into local data experts in their schools (Doll et al., 2010). This reduced the

need for teachers to rely on external expert consultants and strengthened schools' resilience.

Results of the Doll et al. (2010) study showed promise that teachers' data use knowledge, skills, and beliefs improved during participation in the NU Data intervention. In particular, a measure of teachers' data literacy assessed their knowledge of the six Databasics: (1) knowledge of diverse data collection protocols; (2) selecting protocols best suited to answer teachers' questions; (3) collating and graphing data; (4) discerning trends and differences in data; (5) using data in team problem-solving; and (6) selecting evidence-based interventions. Teachers' knowledge of the six DataBasics increased after NU Data training, as did their ratings of the acceptability and likely impact of data-informed instructional modifications. Higher scores for teachers' data use during the project predicted greater student progress towards teacher selected goals. Moreover, pre-, during, and post-examples of teachers' data use showed that their fidelity to standards for high quality data was stronger after completing the NU Data intervention, and these gains were maintained one year after the NU Data Intervention ended (Sikorski et al., 2013). Teachers who participated in the intervention reported having more confidence in their data use skills. Some teachers went on to become data ambassadors in their schools by hosting teacher development programs to teach their staff about the databasics and NU Data intervention.

Purpose of this study

The purpose of this study is to examine the psychometric properties of the NU Data Knowledge Scale, the measure that was developed to assess knowledge outcomes of the NU Data intervention. This dissertation has the following research goals:

1. To assess the overall internal consistency of teachers' scores on the NU Data Knowledge Scale instrument.
2. To assess the capacity of items to discriminate teachers' knowledge of the Databasics; and
3. To describe the descriptive and demographic statistics of the instrument with the sample of teachers.

CHAPTER TWO: LITERATURE REVIEW

The purpose of this study is to examine the psychometric properties of the NU Data Knowledge Scale as a measure of the data literacy of teachers. The previous chapter provided a brief overview of factors in the national educational climate that have increased efforts to measure teachers' data literacy. This chapter will review research related to data literacy, teachers' basic understanding of data, and teachers' data knowledge of the databasics. Additionally, this chapter will describe the databasics that underline the NU Data intervention, the development of the NU Data Knowledge Scale, and components of strong measures.

Data Literacy

The United States Department of Education has invested over 700 million dollars into teacher development programs to increase teachers' knowledge of and familiarity with accessing and interpreting student data at the student, classroom, school, and national level (Lai & Hsiao, 2014). Despite this funding, 72 percent of school districts cited deficiencies in teacher knowledge and skills in data use as barriers to using data effectively in schools (U.S. Department of Education, 2008). As a result, the U.S. Department of Education reinforced the No Child Left Behind Act of 2001's mandate that required teachers to demonstrate deeper understanding of basic assessments and statistical concepts, and fluency in reading and understanding a variety of data representations (i.e., tables, charts, dashboards, database interfaces, and graphing programs; Bakx, Baartman, & van Schilt-Mol, 2014; Means, Chen, DeBarger, & Padilla, 2011; U.S. Department of Education, 2008). Teachers who are fluent in reading, collating, and interpreting student data are better able to apply interventions in response

to the data which can result in better outcomes for students (Christoforidou, Kyriakides, Antoniou, & Creemers, 2014).

Defining Data Literacy. To meet the growing demands of national educational policy, districts sought opportunities to increase the assessment literacy, data-based decision making skills, and teacher knowledge of and familiarity with multiple data use strategies. The term data literacy has been defined many different ways in research, and has often been used as a synonym for assessment literacy, data-based problem solving, data use skills, and formative and summative assessments (Mandinach & Gummer, 2012; Mandinach & Gummer, 2013a; Mandinach, Honey, & Light, 2006; Means et al., 2011; Penuel, Roschelle, & Shechtman, 2007; Stephens et al., 1995; Stiggins, 1995). It has become increasingly important to differentiate data literacy from other forms of data use and teacher knowledge (Mandinach & Gummer, 2013a).

The most widely-quoted definition of data literacy is the ability to understand and use data effectively to inform decisions (Mandinach & Gummer, 2013a). Although concise, this definition does not adequately capture the complexity of data literacy, and has added to the confusion over how data literacy differs from other aspects of data use.

Data literacy is often conflated with assessment literacy in schools because schools heavily emphasize interpreting high-stake assessments (Mandinach, 2012; Mertler & Campbell, 2003, 2005). Assessment literacy examines teachers' abilities to (a) administer and score standardized tests; (b) recognize how their own biases affect interpretation; and (c) use data to guide instructional practices and make educational decisions for students (Jacobs et al., 2009; Mertler & Campbell, 2005). These abilities are also components of data literacy. However, assessment literacy refers to choosing,

selecting, and interpreting large data sets from summative assessments, whereas data literacy incorporates the synthesis multiple measures (i.e., formative assessments, classroom assignments, test scores, etc.) to ecologically examine students' abilities and behaviors (Christoforidou et al., 2014; Quilter & Gallini, 2000).

A key differentiation between data literacy and assessment literacy is data literacy's focus on a wide range of student data like formative assessments (Mandinach & Gummer, 2012), student perceptions (Love, 2011; Vahey, Rafanan, Patton, Swan, van't Hooft, et al., 2012), motivation (Qin & D'Ignazio, 2010; Schield, 2004), and the process of behavior (Mandinach & Gummer, 2013a). The U.S. Department of Education gathered 24 educational experts to define and identify the critical components of data literacy. During this meeting, 76 percent of experts believed that there was overlap between assessment and data literacy; they argued that the major difference from assessment literacy was data literacy's breadth and complexity in application across fields (Mandinach & Gummer, 2012).

Data literacy is also often confused with data-based decision making. Like data literacy, data-based decision making emerged out of federal policy changes and legislation that encouraged teachers and researchers to examine how data impacts educational decisions for students (Dunn, Airola, Lo, & Garrison, 2013). Federal legislation emphasized incorporating comprehensive student data into assessments instead of relying on only standardized assessments (Dunn et al., 2013; Newton, Algozzine, Algozzine, Horner, & Todd, 2011; Roehrig, Duggar, Moats, Glover, & Mincey, 2008). As a result, schools needed to broaden their assessment practices to include aspects of behavior and motivation (Duncan, 2009). The No Child Left Behind

Act also mandated that schools show that students make Adequate Yearly Progress (AYP) on high stakes assessments this mandate further pushed schools to provide teachers trainings and instruction in data use (Wayman et al., 2013).

To meet the new federal criteria for AYP and proficiency standards, districts incorporated data-based decision making practices into their classrooms. According to the National Association of School Psychologists (NASP) best practices in data-based decision making practices require teachers to (Dunn et al., 2013; Ysseldyke et al., 2009): (1) conduct comprehensive evaluations of students that incorporate observations, interviews, and standardized assessments; (2) implement curriculum-based assessments, measurements, and evaluations to monitor student progress; (3) employ ecological assessments to measure the impact of the school environment on learning; and (4) differentiate instruction to meet the specific educational needs of students.

These four components are critical to understanding both data-based decision making and data literacy. Like data-based decision making, data literacy requires using a variety of data sources to make educational decisions for students (Jacobs, Gregory, Hoppey, & Yendol-Hoppey, 2009). Moreover, one aspect of data literacy is engaging in independent and collaborative team problem solving that translates student data into actionable practices (Mandinach & Gummer, 2013a). An important difference between data literacy and data-based problem solving is that data literacy focuses more broadly on basic teacher knowledge of data collection practices and skills; whereas data-based problem solving focuses more on the iterative process of data collection and hypothesis testing (Huguet, Marsh, & Farrell, 2014; Wayman & Jimerson, 2014). In this regard, data

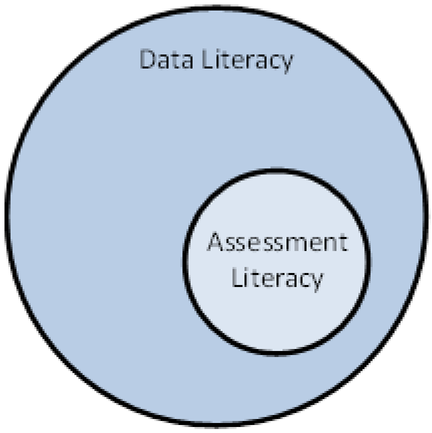
literacy looks beyond how teachers use student data and focuses on how teachers use student data well (Jimerson, 2014).

Data literacy is consistent with the Response to Interventions (RtI) framework for using student data to identify students for special educational services (Fuchs & Fuchs, 2006). The goal of RtI is to describe students' eligibility for special education by using curriculum based measures to monitor their progress in the classroom while implementing responsive classroom instruction that is sensitive to specific student needs (Ingram, Louis, & Schroeder, 2004). Likewise, data literate teachers are empowered to use student data in their classrooms to drive instruction, and to change educational practices to improve the educational outcomes of all students (Stephens et al., 1995). Indeed, RtI reflects many positive qualities of data literacy (i.e., actionable practices, progress monitoring, and changing educational practices in response to student data), and is an integral component of the data literacy framework.

In essence, data literacy may be thought of as the next iteration of assessment literacy, data-based decision making and RtI. Multiple illustrations have been attempted to distinguish data literacy from other forms of data use. One illustration that emerged from the U.S. Department of Education (2011) summit on data literacy was a Venn diagram (see Figure 1:1) in which assessment literacy is shown as housed within data literacy.

Figure 1:1

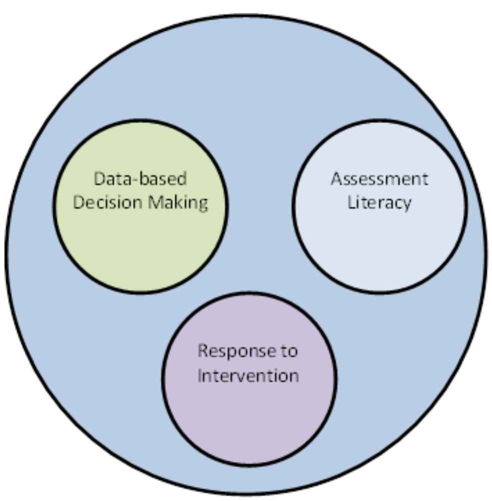
Venn diagram from (Mandinach & Gummer, 2013)



A majority of data use experts at the summit postulated that data literacy possessed all of the aspects of assessment literacy, and also had its own characteristics. Alternatively, data literacy can be described in a Venn diagram as comprising assessment literacy, data-based decision making, and actionable data use practices (see Figure 1:2). However, these diagrams do not go far enough in identifying the critical components of data literacy.

Figure 1:2

Revised Venn Diagram of Data Literacy



Components of Data Literacy. In order to be data literate, teachers need to have basic knowledge of data (Vahey, Rafanan, Patton, Swan, van't Hooft, et al., 2012), to apply and interpret sound data use techniques and information (Newton et al., 2011), and use the data they collect to guide future educational decisions. This review will describe three distinctive frameworks describing data literacy: The Mandinach and Gummer (2012) framework, the Means et al. (2011) framework, and the NU Data (Doll, 2010) framework.

There is growing consensus about which critical components of data literacy are most salient (Dunn et al., 2013). Mandinach and Gummer's (2012) framework postulated five critical components of data literacy: data location, data interpretation, data comprehension, question posing, and data use. These components were derived from conducting an extensive literature review of data literacy, data-based decision making, and teacher data use. In total, 41 books and 6 articles were reviewed and independently coded by Mandinach and colleagues (2012) until they arrived at the five outlined components. They defined data location as the ability to find the right kinds of data to answer specific questions. Data interpretation was defined as the ability to derive meaning from the data, and data comprehension was defined as knowing what the data say. Question posing was defined as the ability to use data to answer future questions, and data use was the ability to base future instruction, interventions, and planning based on student data. These components supported national efforts to support effective teacher data use in schools (Mandinach & Gummer, 2012).

Similarly, the Means et al. (2011) framework identified four content domains comprising data literacy. These included: Data Knowledge, content knowledge and

teacher; Data Focus, ability to access a variety of data protocols and assessment techniques; Problem Focus, ability to identify problem and formulate hypotheses; and Process Focus, collaborative inquiry, knowing how to solve data use problems (Means et al., 2011). These components were developed through consensus building strategies with 76 experts. Means and colleagues (2011) developed these components to address the missing emphasis on teacher data skills and data informed decision making in previous data literacy research.

Both the Mandinach and Gummer (2012) and Means and colleagues (2011) frameworks align across four broad domains. These domains are: (1) teachers need to have basic knowledge and skills in data use, (2) teachers need to be able to interpret the data they collect, (3) data needs to be used in classroom settings, (4) teachers need to be able to engage in data based problem solving.

Teachers need to have basic knowledge and skills in data use. The first broad component of data literacy describes foundational skills for teachers use and manipulation of basic forms of data (Means et al., 2011). Teachers need to be able to differentiate between measures of central tendency (i.e., mean, median, and mode) and understand how they are used (Dunn et al., 2013; Ingram, Louis, & Schroeder, 2004). They need to understand and be able to interpret graphs and other basic forms of student data. In educational settings, teachers must be able to understand, create, and manipulate data in graphs in order to understand student performance on progress monitoring curriculum-based measure and standardized assessments (Jacobs et al., 2009; Mandinach & Gummer, 2012; McCutchen et al., 2002).

This component represents the foundational knowledge of data use that teachers demonstrate throughout their careers. It is their familiarity with interpreting student grades, creating rubrics for assignments, and collecting behavioral and academic intervention data on students (Roehrig et al., 2008). This foundational knowledge is predicated on teachers' pre-service trainings and experiences in the classroom. Preliminary results from the NU Data intervention showed that teachers had some basic knowledge in data use, such as: knowing how to collect data on student behavior, academic progress, and efficacy; and teachers had a basic, but insufficient, knowledge of the databasics and data use strategies (Sikorski et al., 2013). Basic teacher knowledge in data use is essential for teachers to be able to use data effectively to increase the educational outcomes for students.

Teachers need to be able to interpret data they collect. The second domain of data literacy, data interpretation is defined as the teacher's ability to derive meaning from student data, protocols, and measures (Mandinach, 2012; Mandinach & Gummer, 2013a). Data interpretation skills are important for teachers to be able to draw qualitative and quantitative inferences from student data such as classroom observations, descriptions of student behavior, and hypothesis testing. Teachers who can interpret results from a variety of assessments, behavioral, academic, and formative measures are highly skilled (Means et al., 2011).

Preliminary results from the NU Data Intervention study (Doll et al., 2014) showed that teachers were missing critical skills in organizing and interpreting the data they collected on students prior to completing the NU Data intervention. After the

intervention, teachers were more skilled in assessing student data, but lacked the skill and knowledge to be able to apply interventions in response to student data.

Data needs to be applied in classroom settings. The third domain of data literacy is data application. Data application describes how well teachers apply data use skills within classroom settings. This is evident by teachers turning data into actionable practices. An example of data application would be a teacher adjusting their curriculum in response to student scores on a benchmark assessment. Teachers who are able to apply data use skills within the classroom are better able to differentiate instruction in response to student data to meet specific student needs (McCutchen et al., 2002; Roehrig et al., 2008). Data application skills are integral to RtI in which teachers monitor, track, graph, and interpret formative assessments, benchmark scores, and curriculum-based measures. Additionally, data application incorporates components of assessment literacy because it includes teachers' abilities to incorporate summative assessment results into curriculum planning, teaching pedagogy, and instruction (Jacobs et al., 2009; Lukin, Bandalos, Eckhout, & Mickelson, 2004; Smith, 2013).

Teachers also need to able to gather and synthesize multiple forms of data when assessing student progress (Mandinach, 2012; Mandinach & Gummer, 2013b; Roehrig et al., 2008; Smith, 2013). This ensures that adjustments made to student curriculum are based on a comprehensive review of student data from multiple sources, rather than from a single measure. Comprehensive evaluations synthesize data collected from direct and indirect sources, including: student records, grades, behavioral observations, formative and summative assessments, curriculum-based measures, rating forms, interviews, and

high stakes testing (Mertler & Campbell, 2003; Smith, 2013; Stephens et al., 1995; Ysseldyke et al., 2009).

Teachers need to be able to engage in Data-based problem solving. The fourth and final component of data literacy is data-based problem solving. Data-wise teachers are able to formulate hypotheses based off student data, collect meaningful baseline data, and change instructional techniques to meet individual student needs (Boudett, City, & Murnane, 2005). In RtI, data literate teachers use data-based problem solving strategies to set goals, monitor progress, and evaluate student outcomes. Data-based problem solving strategies incorporate continuous assessment strategies and iterative problem-solving techniques to answer hypotheses about students. Teachers participating in data-based problem solving strategies often collaborate with other professionals to examine data and dialogue (Newton et al., 2011). Because data literate teachers are experts in the use of student data, they are able to describe data to parents and other teachers in an easily understood way (Doll et al., 2005).

The Mandinach and Gummer (2012) and Means et al. (2011) frameworks adequately identified most of the critical components of data literacy. However, these critical components are not exhaustive and do not address some of the barriers to measuring and identifying data literacy. Additionally, researchers do not agree which data components stand alone and which components are interrelated with each other (Mandinach & Gummer, 2013b). Consequently, there are some gaps in published descriptions of the critical components of data literacy.

The first critical component not identified by these two definitions is the impact and importance of teacher training and development programs. Indeed, teacher

experiences are critical to developing the data literacy and data-based decision making ability of teachers (Love, 2004, 2011). However, there are no current educational standards in pre-service teacher courses that target data use skills (Jimerson, 2014; Staman, Visscher, & Luyten, 2014). As a result, teachers' mastery of data literacy is inconsistent when teachers first enter the workforce (Mandinach & Gummer, 2013b). This inconsistency results in teachers having different needs in data use trainings throughout their careers. The Council for Accreditation of Education Preparation (CAEP; Council for the Accreditation of Educator Preparation, 2013) mandated that teachers be able to use valid assessments and provide a rationale for using them in the classroom. However, this does not require teachers be data literate; rather it mandated programs to teach teachers to be articulate about assessment.

School districts have spent millions of dollars on professional development courses in data literacy (Jimerson, 2014; Lai & Hsiao, 2014). Unfortunately, these trainings are quite variable: 90 percent of school districts' professional development programs are provided to only a fraction of the school staff (Mandinach & Gummer, 2013b). This reinforces differential exposure to and attitudes towards data use in schools, and detracts from a culture of broad and shared responsibility for data (Jimerson, 2014). When requirements of teacher training programs are not systematic, teachers will not use consistent language when discussing student data, will implement different classroom behavioral and academic interventions, and staff-wide buy-in of intervention techniques is difficult (Staman et al., 2014).

A second limitation of the components of data literacy identified by the Mandinach and Gummer (2012) framework and Means et al. (2011) framework, is that

neither framework emphasized the importance of fostering cultures of support for data use (Jacobs et al., 2009; Jimerson, 2014). Schools create systematic support for data literacy among their staff by providing staff-wide trainings in data use and support (Love, 2004, 2011; Mandinach & Gummer, 2013b). There is a growing need for teacher training and preparation programs that schools could use to increase data literacy knowledge and skills of their teachers**.** This would provide teachers with a shared reference point when discussing student data and engaging in team-based problem solving.

DataBasics

An alternative framework of data literacy has emerged out of teacher consultation research conducted by Doll and colleagues (2005). The databasics are early and very basic data skills that provide critical foundations for data literacy as described by Mandinach and Gummer (2012) and Means et al. (2011). The databasics were developed by Doll et al. (2005) through a mixed-method study examining barriers that teachers faced when trying to implement high quality data use procedures with fidelity. The study conducted quantitative analyses of teacher surveys and qualitative analyses of focus groups transcripts from 76 teachers in 13 schools.

Analyses identified six strategies and skills that teachers believed would increase their data-based decision making skills. Doll et al. (2010) referred to these skills as databasics: (1) knowledge of diverse data collection protocols; (2) selecting protocols best suited to answer teachers' questions; (3) collating and graphing data; (4) discerning trends and differences in data; (5) using data in team problem-solving; and (6) selecting evidence-based interventions. Though developed as part of consultation research, Table

1:1 shows how the databasics align with the critical components of data literacy described by the Mandinach and Gummer (2013) and Means et al. (2011) frameworks.

Table 1:1

NU Data Pragmatics Compared to Critical Compoents of data Literacy

Doll, Horn, Shope (2010)	Mandinach and Gummer (2012)	Means et al. (2011)
1. Knowledge of data collection protocols	*Data location*, finding the right kinds of data to answer specific questions	*Data Knowledge*, content knowledge and teacher pedagogy *Data Focus*, knowing how to access a variety of data protocols and assessment techniques
2. Using protocols to best suited for team's questions		*Problem Focus*, abilty to identify problem and formulate hypotheses
3. Collating and Graphing data	*Data Comprehension*, knowing what the data say	
4. Discerning trends and differences in data	*Data Interpretation*, deriving meaning from the data	
5. Team problem solving model	*Question posing*, using data to guide future questions	*Process focus*, collaborative inquiriy,

	Data use, adjusting future instruction, interventions, and planning based off data	knowing how to solve data use problems
6. Selecting evidence-based interventions		

Further evidence of the relevance of the databasics was found in a subsequent study. Kosse (2006) examined the relation between the problem solving strategies used by student assistance teams and the resulting student outcomes. Results demonstrated that teachers implemented the problem solving procedures with greater fidelity when provided with explicit tools for gathering, analyzing, and using data on student learning (Kosse & Doll, 2006). Subsequent studies have similarly shown that teachers often believe they lack access to resources or time to implement classroom-based interventions well and to collect student data (Ingram et al., 2004; Roehrig et al., 2008). The databasics framework is consistent with the critical components of data literacy outlined by Mandinach and Gummer (2012) and Means et al. (2011) while addressing some of the gaps in the literature outlined earlier.

DataBasic 1: Knowledge of a variety of data collection tools. The first DataBasic corresponds to Mandinach and Gummer's (2013) data location and Means et al. (2011) data knowledge and data focus components. This databasic describes teachers' basic knowledge of multiple data collection use skills and protocols (Sikorski et al., 2013). There is an old saying that when your only solution is a hammer, every problem begins to look like a nail (Maslow, 1962). Applied to data-based problem solving, if teachers only know one or two data collection tools, these will be the only data they use

to make educational decisions for students. In schools, mastery of DataBasic 1 is demonstrated when teachers have in-depth knowledge of multiple alternative strategies to use for collecting a wide variety of student data (Bernhardt, 2013; Goldring & Berends, 2008; Jacobs et al., 2009; Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006). Teachers need familiarity with data collection tools for tracking the academic achievement, behavioral success, and academic engagement of their students (Abbott, 2008; Anderson, Leithwood, & Strauss, 2010; Bernhardt, 2013; Brunner et al., 2005; Love, 2011).

Since there are wide discrepancies in teacher pre-service education and in-service teacher development programs, many teachers do not have the necessary skills to use data effectively in their classrooms (Choppin, 2002; Mandinach, 2012). As a result, teachers have wide variances in their knowledge, skills, and beliefs of data use, and may not have basic knowledge of effective data use strategies (Doll et al., 2010). Despite differences in their pre-service trainings, however, teachers develop some data use expertise as they gain work experience (Love, 2004). As teachers develop expertise throughout their careers, they cultivate preferences for certain data collection protocols. Data literate teachers use protocols that are brief and easy to administer (Roehrig et al., 2008), use protocols that easily translate to graphing software (McCutchen et al., 2002), are easily converted to graphical display (Boudett et al., 2005), and will effectively answer the questions that teachers are trying to answer (Christoforidou et al., 2014). They have knowledge of basic data use strategies and protocols that are strengthened as they gain more work experience.

DataBasic 2: Selecting data collection tools that fit teachers' questions about students. The second DataBasic describes the match between data collection strategies

and the questions teachers are trying to answer. In order to increase the educational outcomes of students, teachers need to have the knowledge and skills to determine which types of protocols are needed to evaluate students' academic efficacy, behavioral success, and academic achievement (Booher-Jennings, 2005; Brunner et al., 2005; Love, 2004; Mandinach & Gummer, 2012; Reschly & Christenson, 2006; Sikorski et al., 2013). Data literate teachers can differentiate the levels of data (e.g., student or classroom level) that are most useful in answering their questions about students (Brunner et al., 2005; Love, Stiles, Mundry, & DiRanna, 2008). They select protocols that address their target behavior at the student, cluster of students, class, grade, or school level.

In practice, teachers' ability to select the proper protocols to answer their hypotheses is limited by their training and work experiences (Means et al., 2011). This second DataBasic was illustrated in the NU Data intervention through the use of case examples, individual and team-based coaching, and feedback to help teachers notice the match between interventions and hypotheses (Doll et al., 2013). Through weekly coaching sessions, teachers were able to expand their knowledge of how to select and apply evidence-based interventions in their classroom and had more success addressing the needs of their students (Sikorski et al., 2013).

DataBasic 3: Collating and graphing data. This DataBasic corresponds to Mandinach and Gummer's (2012) critical component of data interpretation. Data literate teachers are able to collate data from multiple protocols and assessments into easily readable graphs by hand or using spreadsheets (e.g., Excel; Christoforidou et al., 2014; Lai & Hsiao, 2014; McCutchen et al., 2002; Staman et al., 2014). Properly graphed data allows teachers to visually analyze student data and determine the means, trends, and

levels from the graphical output (Huguet et al., 2014; Jimerson, 2014; Wayman & Jimerson, 2014). This databasic is imperative because well organized and graphed student data allows teachers to easily share data with parents and other professionals and describes student progress towards goals (Kazdin, 2011).

In practice teachers use graphs for to accomplish a number of professional tasks (Bautista, Canadas, Bizuela, & Schlimann 2015). Teachers use graphs to monitor student progress towards goals in RtI assessments, chart student behaviors, and view student progress on national assessments (Huguet & Farrell, 2014). Teachers need to be able transcribe raw student data into visual representations that allow them to determine trends and differences in student outcomes (Wyman & Jefferson, 2005).

DataBasic 4: Interpreting student data. The fourth Databasic corresponds to Mandinach and Gummer’s (2012) component of data comprehension. Data literate teachers are able to interpret data from multiple sources down to the individual item-level (Boudett et al., 2005; Jacobs et al., 2009; Jimerson, 2014). They are able to interpret data from multiple levels, including: high stakes assessments, formative assessments (i.e., RtI data), classroom assignments, and behavioral protocols (Lai & Hsiao, 2014; Mandinach & Gummer, 2013b; Staman et al., 2014; Suen, Lei, & Li, 2011). This means that data literate teachers are able to determine what the data mean in the proper contexts (Roehrig et al., 2008). This allows them to draw inferences from the data they collect.

In practice, teachers need to be skilled in interpreting raw student data and visual representations like graphs (Wyman & Jefferson, 2005). Strong data interpretation skills allow teachers to monitor student progress and student achievement by reviewing the trends, variability, and levels in student data (Doll et al., 2014). This is vital in

determining whether interventions or teaching strategies are being effective (Means et al., 2011).

DataBasic 5: Using data to refine instructional modifications. The fifth DataBasic corresponds to Mandinach and Gummer's (2012) component of question posing and Means and colleague's (2011) component of problem focus. This component of data literacy describes teachers' ability to turn student data into actionable teaching practices that fluently adjust curriculum or student interventions in response to data (Jimerson, 2014; Kerr et al., 2006). Data-savvy teachers are able to use the results of student data to guide instruction and refine the curriculum (Vahey et al., 2012).

In practice, data literate teachers are able to collaborate with teams and engage in problem solving (Gulikers, Biemans, Wesselink, & van der Wel, 2013; Halverson, Prichett, & Watson, 2007; Long, Rivas, Light, & Mandinach, 2008). When data is being used to drive instruction, data literate teachers look for patterns in student data to refine interventions as a result of collaborating with other teachers (Boudett et al., 2005).

DataBasic 6: Selecting evidence-based interventions. The sixth DataBasic describes teachers' abilities to find high quality evidence-based practices that match their intervention goals. Data literate teachers are familiar with and know how to access interventions review sites such as: What Works Clearinghouse and interventioncentral.org. Additionally, these teachers are able to choose evidence-based interventions that match their classrooms' demographics, needs identified in their data, and culture with the understanding that interventions need to be adjusted for specific school contexts (Gettinger, Mulford, & Hoffman, 2010).

Taken together, these six databasics describe the framework for data literacy that underlay this dissertation's NU data Knowledge Scale. The benefit of the six databasics is that they emerged out of teacher consultation research and describe barriers teachers face in schools.

Barriers to assessing data literacy

Previous reviews of data literacy have called for the development and distribution of a brief, multiple choice measure of data literacy with strong psychometric properties (Jacobs et al., 2009; Mandinach & Gummer, 2012; Means et al., 2011; Schoenfeld, 2007). A search of What Works Clearinghouse, Google Scholar, EBSCO, SAMSHA, and PsycInfo revealed no empirically validated measures of data literacy. Mandinach and colleagues (2012) and the Means et al. (2011) called for the development and validation of a brief, multiple-choice assessment of teacher's knowledge of data literacy and data use skills. Since no empirically-validated measure of data literacy is currently described in the published literature, districts have developed their own measures of data use and data-based decision making (Ingram et al., 2004). These measures are limited in scope and typically are designed for a specific trainings or programs.

What makes a useful measure? A strong measure of data literacy should be a 20 to 30 item multiple-choice measure (Means et al., 2011). Compared to scenario-based short answer or essay measures, multiple-choice instruments are easier to administer, score, and interpret (Thorndike & Thorndike-Christ, 2010); are cheaper to proctor, and are less time consuming to administer than other response formats (Haladyna, Downing, & Rodriguez, 2002; Haladyna & Rodriguez, 2013). One criticism of multiple choice measures is that there is probability of random guessing. However, the probability of

someone guessing ten out of ten answers correctly on a ten-item test is less than .0000001 percent (Haladyna & Rodriguez, 2013).

The best multiple-choice items are based on scenarios or examples that are relevant to working in the classroom (Frey, Petersen, Edwards, Pedrotti, & Peyton, 2005; Haladyna et al., 2002; Haladyna & Rodriguez, 2013). Thus, it would be beneficial for multiple-choice items on a data literacy measure to be developed and modeled based off of teacher experiences.

Means and colleagues (2011) recommended that test items be formed from item banks of short answer questions so that the stems and distractors appear genuine (Frey et al., 2005; Haladyna et al., 2002; Hansen & Dexter, 1997; Means et al., 2011). High quality test items should be discriminative. Item-total correlations for individual items should be above .2 showing that the item discriminates teachers by their knowledge (Haladyna et al., 2002; Haladyna & Rodriguez, 2013). Individual items should also have discriminative distractors with negative point-biserial correlations for incorrect distractors with the overall score. The correct answer should have a positive point-biserial correlations with the overall score, and should have point-biserial correlations greater than .2 (Frey et al., 2005; Haladyna et al., 2002; Haladyna & Rodriguez, 2013).

Strong multiple-choice measures should also be reliable and valid. Kuder-Richardson 20 is a recognized method for estimating a measure’s internal consistency and represents each item as a one-item exam that is correlated with the total test scores (Thorndike & Thorndike-Christ, 2010). KR-20 provides the internal consistency in measures that are scored dichotomously right or wrong. A Kuder-Richardson 20 score of

at least .7 or above is indicative of a highly consistent instrument measuring a single latent variable such as data literacy (Knapp & Mueller, 2010; Suen et al., 2011).

A strong measure of data literacy should also have a factor structure that aligns with the underlining conceptual framework (Bakx et al., 2014; Means et al., 2011). Conceptually consistent factor structures provide construct validity evidence that is crucial when describing the dimensionality of the scale (Costello & Osborne, 2005). Measures being developed should be examined with an exploratory factor analysis using the target population of interest (Worthington & Whittaker, 2006). This requires a population of at least 100 teachers or a ratio of five people per test item (whichever is larger) in order to have sufficient power to conduct the analyses (Gorsuch, 1997). When conducting weighted least means squares to explore factors, it is critical to determine what the markers will be for determining which factors to retain (Hayton, Allen, & Scarpello, 2004; Schaughency & Ervin, 2006; Worthington & Whittaker, 2006).

When determining which factors to retain, it is useful to conduct parallel analyses by conducting a series of principal component analyses of randomly generated data and taking the mean and 95th percentile of all observed eigenvalues (Hayton et al., 2004). These eigenvalues are then plotted on top of the true data to create cutoff points for retaining factors. If the eigenvalues from the actual data are greater than the eigenvalues from the parallel analyses, then they are retained. Fit indices of the exploratory factor analysis should have an insignificant chi-square, “goodness of fit” (GFI) above .9, comparative fit index (CFI) greater than .9 (Marsh, Balla, & McDonald, 1988), a Tucker-Lewis Index (TLI) above .95 (Haladyna & Rodriguez, 2013), and a standardized root mean square (SRMS) below .05 (Steiger & Lind, 1980).

Finally, a strong measure should provide evidence of content validity (Cronbach & Meehl, 1955). In this study, content validity is recognized as providing evidence of agreement between subject matter expert categorizations of the test items by databasics. Subject matter experts are people who can provide significant, in-depth analysis of the content of a measure (Wynd, Schmidt, & Schaefer, 2003). Subject-matter experts are specialists holding expertise within a specified field related to the measure. They provide expert ratings of each item's relevance and make a judgment of whether the collection of items adequately represents the content domain being tested (Knapp & Mueller, 2010; Lawshe, 1975; Sireci, 1998). When measures are strong, expert ratings of an item will match 60 percent of the time (Thorndike & Thorndike-Christ, 2010). Since percentage of agreement is distorted when events are low incidence, a more useful measure of interobserver agreement is kappa. Kappa estimates reflect the difference between the observed agreement and chance agreement (Thorndike & Thorndike-Christ, 2010; Viera & Garrett, 2005). Kappa scores can range from -1 to 1 with 1 being perfect agreement, 0 being the chance agreement, and -1 indicating less than chance agreement (Viera & Garrett, 2005; Wynd et al., 2003). When agreement is based off of non-likert ratings, then a non-weighted kappa can be used (Viera & Garrett, 2005). The kappa estimate is useful in describing the level of agreement with the goal of achieving at lease a moderate level (.41) of agreement.

Development of the NU Data Knowledge Scale

The NU Data Knowledge Scale was developed as part of a three year, IES-funded, research study of teacher data use practices (the NU Data Intervention Study). At the beginning of the first year of the study, a panel of researchers and data experts

developed five to seven short-answer items for each of the six databasics. From this item pool, two open ended items were selected for each databasic to construct a 12-item short-answer measure. This measure was administered to 17 teachers. The internal consistency of the measure was low, so items were revised and the total number of items was increased to 20.

Three months later, the refined 20-item measure was completed by the same 17 teachers and they were asked to provide feedback on item wording and item difficulty. Responses were scored and coded by two independent research assistants using a 0 to 2 scale. Responses that showed in-depth understanding of the content were scored a 2; responses that reflected basic understanding of the answers were scored a 1; and responses that were incorrect or missing were scored a 0. The interclass correlation (ICC) of absolute agreement between the two raters was high with ICC values for absolute agreement of .94. The feedback provided by the teachers was used to refine the items' wording and to increase the perceived validity of the items for teachers. Results showed that the 20-item measure had excellent internal consistency with a coefficient alpha score of .95. Despite the high coefficient alpha, three items were removed based on teacher recommendations, shortening the measure to 17-item short-answer items.

Two months later, the 17-item measure was completed by the same 17 participants. Responses were again scored using a three point likert scale: 0 for an incomplete, missing, or incorrect answer, 1 for an answer that reflected basic understanding, and 2 for an answer that reflected enriched understanding. Again the inter-rater reliability of absolute agreement between the two independent research assistants was high with an ICC score of absolute agreement of .815.

Next, the short answer responses from the 17-item measure were used to create an item bank of multiple-choice items assessing the six databasics. Common incorrect responses were used to generate quality distractors for the test stems. Parallel items were developed from teacher responses. The result was a 20-item multiple-choice measure of teacher knowledge related to the six databasics.

A content analysis of the 20-item multiple-choice measure was completed by four data experts who identified which databasic were assessed each item. This content analysis was the basis of a likely factor structure of NU Data Knowledge Scale. It was hypothesized that the NU Data Knowledge Scale had a single-factor construct of data literacy given the overlap that experts perceived between the six databasics.

Ten additional multiple-choice items were then developed from the item bank to construct a 30-item multiple-choice measure. Revisions were made to the items to remove confusing distractors, rephrase question stems, and adjust images (e.g., graphs embedded within the measure) to be more clear and concise. The 30-item measure was completed by 17 teachers participating in Year Two of the NU Data Intervention study in October, December, and April. Teachers were given two copies of NU Data Knowledge Scale during the October meeting (one on white paper and one on yellow paper). They were instructed to complete the NU Data Knowledge Scale on the white paper version and provide feedback and revision on the yellow version. The feedback from the yellow version of the measure were used to incorporate the teachers' expertise in future revision of the measure. Additional feedback was collected from data experts, and results of analyses prompted revisions to the stem and distractors of several items of the revised 30-item NU Data Knowledge Scale.

The revised 30-item measure was then piloted with 168 pre-education undergraduate students at a large Midwestern university to examine the properties of the measure. Results suggested that the scale had acceptable internal consistency ($\alpha = .763$) and showed evidence of being a single factor construct ($\chi^2 = 1231.474$, $df = 435$, $p < .0001$; CFI = .971, TLI = .968, RMSEA = .019). Analyses of the items and distractors resulted in seven items being flagged because they had low item-total correlations or had distractors that were positively correlated with the total score. As a result, adjustments were made to these items, and the scale was then converted to an electronic survey.

The 30-item electronic measure was administered to 16 Year Three teachers in September, December, and March. The measure demonstrated an acceptable level of internal consistency with a coefficient alpha score of .85. Analysis of the sensitivity of the NU Data Knowledge Scale to measure growth in teacher knowledge of the databasics showed that Year Three teachers' scores on the NU Data Knowledge Scale increased over the duration of the intervention. Across the three years of the study, teachers' scores were significantly higher on the NU Data Knowledge Scale at the end of the study year than the beginning (Estimate = 1.22; SE = .58; $p = .040$).

A limitation of the 30-item NU Data Knowledge Scale in its present form is that the EFA and item analyses, although promising, were calculated using data gathered from 168 pre-service students. Additional research is needed to determine the effectiveness of the 30-item NU Data Knowledge Scale with a larger sample of in-service teachers. Additionally, the current measure's content validity should be examined once again by a team of experts. With additional validity support from subject matter experts and

implementation of the measure with a large teacher sample, this measure could be used to support national efforts to strengthen teachers' data use.

Research Goals

The purpose of this study is to examine the technical soundness of the NU Data Knowledge Scale in measuring teachers' data literacy. This study has three research goals with six research objectives:

Research Goal 1: To assess the overall internal consistency of teachers' scores on the NU Data Knowledge Scale.

a. To describe the descriptive statistics of the instrument for the sample of teachers; and
b. To assess the overall consistency of teachers' scores on the instrument

Research Goal 2: To assess the capacity of items to discriminate teachers' knowledge of the DataBasics.

a. To examine the alignment of the subject matter experts' ratings with the test blueprint of the NU Data Knowledge Scale; and
b. To determine the dimensionality of the scale through weighted least mean squares of mean variance (WLSMV).

Research Goal 3: To describe the relation between performance on the NU Data Knowledge Scale and teachers' characteristics.

a. To describe the degree to which years of experience, gender, classroom size, Comfort with Data Use scale scores, or teaching specialization predict scores on the knowledge scale.

CHAPTER THREE: METHOD

Research Design

This study has three research goals with six research objectives. The primary goal of this study is to perform a psychometric analysis of the NU Data Knowledge Scale using a sample of teachers. This primary goal of this study is to assess the overall internal consistency of teachers' scores on the NU Data Knowledge Scale by: (1) to describe the descriptive statistics of the instrument for the sample of teachers; and (2) assessing the overall consistency of teachers' scores on the instrument.

The second goal of this study is to assess the capacity of items to discriminate teachers' knowledge of the databasics by examining the alignment of the subject matter experts' categorizations of the NU Data Knowledge Scale by databasic(s); and (5) determining the dimensionality of the scale through weighted least mean squares of variance analyses.

The third goal of this study is to describe the relation between performance on the NU Data Knowledge Scale and teachers' characteristics. This final goal will answer the following research objective: (6) to describe the degree to which years of experience, gender, classroom size, teaching assignment, teaching specialization, or Comfort with Data Use scores predict scores on the knowledge scale.

Participants

Teachers from Nebraska Schools. Two-hundred and three principals were emailed to ask for their consent to contact their teachers. Teachers in this study were recruited from 39 schools across Nebraska that consented to participate for a school participation rate of 19.2 percent. Of the 215 participants, 177 were female and 38 were

male. The proportion of male teachers 17.67 percent is below the national average of 25 percent (U.S. Department of Education, 2013). Participants in this study were diverse in their years of experience, highest degree received, specialization, and teaching assignments. Overall, participants were highly educated with over half (66.94) having received at least a Master's degree in education. These percentages are higher than the state's average of 50.24 percent of teachers having a Bachelor's or Associate's degree, and 49.6 percent of teachers having a Masters, EdS, or PhD (U.S. Department of Education, 2013). Likewise, participants in this study were more apt to have over 11 years of experience (53.02 percent) than 10 years or less of teaching experience (46.98 percent).This was similar to the state averages of 37.8 percent of teachers having 10 or less years of experience, and 62.2 percent having 10 or more years of experience (U.S. Department of Education, 2013). Teachers in this study had more students in their classroom than the state average (50 percent) with 66.82 percent of teachers reporting having class sizes over 21 students (U.S. Department of Education, 2013). Finally, teachers in this study were less ethnically diverse than the national average with 97.21 percent of teachers identify as White (non-Hispanic) or Caucasian compared to 84.32 percent of the National average (U.S. Department of Education, 2013). All demographic data describing the teacher participants are displayed in Table 3.1.

Table 3.1

Teacher Demographic Information

Characteristic	Teachers (n)	Percentage
Gender		
Male	38	17.67
Female	177	82.32
Ethnicity		
White or Caucasian	209	97.21
Black or African American	0	0
Asian or Pacific Islander	0	0
Latino or Hispanic	5	2.32
American Indian or Alaskan Native	0	0
Other	1	.47
Highest Degree		
Associates	1	.005
Bachelor's	67	31.12
Masters	110	51.12
EdS/MA+30	34	15.81
PhD	3	.01
Years of Experience		
1 to 5	43	20.00

6 to 10	58	26.98
11 to 15	32	14.88
16+	82	38.14
Specialization		
General Education	154	71.63
Special Education	25	11.63
Speech/Language	4	1.86
Other	32	14.88
Subject(s)		
General Ed. Elem.	77	35.81
General Ed. Secondary	14	6.51
Secondary Math	23	10.70
Secondary Science	13	6.05
Sec. Soc. Studies	10	4.65
Secondary English	18	8.37
Other	79	36.74
Number of students		
0-4	7	3.24
5-10	18	8.33
11-15	17	7.79
16-20	43	19.91
21-25	72	33.33
26-30	30	13.89

31+ (rotating)	28	12.96

Nebraska was an ideal location for this study because of the state's disproportionate gaps in achievement between students from middle class families (whose achievement ranks among the highest in the nation) and students from low income families. Nebraska also has the fourth highest Black-White achievement gap and the fifth highest achievement gap for fourth grade students (National Assessment of Educational Progress, 2013). Nebraska also has several districts with minority-majority student populations where over half of the students speak English as a second language. As a result, teachers and administrators rely heavily on the use of classroom and district level data to guide instructional practices (Jacobs et al., 2009).

A review of literature was conducted to determine an adequate sample size for this study. A sample size of at least 150 teachers is suggested to calculate the Kuder-Richardson 20 and item-total correlations for a 30-item measure (Keppel & Wickens, 2004). There is some disagreement in the literature as to what constitutes an adequate sample for determining the dimensionality of the scale. Comrey and Lee (2013) determined sample sizes of 100 as poor, 200 as fair, 300 as good, 500 as very good, and 1000 or more as excellent. However, others have said that the best practice in factor analysis is using a 5:1 ratio of teachers to items analyzed (Thompson, 2004). This study's sample of 215 teachers is above the minimum requirement of 150 participants to conduct the analyses and within the "fair" range for overall sample sizes.

A post hoc analysis of power was conducted to determine whether there were enough participants to calculate the root mean square (Preacher & Kaufman, 2006). The calculation of statistical power determined that there was a power of 1 for this study. This

means that there was adequate power to detect differences between the null model and single factor solution.

Subject Matter Experts. Participants also included a three-member panel of educational subject matter experts who independently reviewed and coded the test items by databasics. Three subject matter experts were the optimal number for this study as it avoided practical limitations of scheduling conflicts and the odd number of experts prevented a tie-vote during the consensus building process. Subject matter experts provided expert judgment of the measure’s wording for stems, selection of distractors, and alignment to the databasics conceptual framework. The subject matter experts from this panel were nationally recognized as experts in data use as evidenced by their contributions to the field and years of experience. Three Subject Matter Experts were recruited by the principal investigator and were provided a description of the study, consent form, copy of the NU Data Knowledge Scale, and Databasics Matrix. Once consent was obtained, the principal investigator instructed the subject matter experts to review the NU Data Knowledge Scale and to make notes as to which items or distractors might need revisions to increase item clarity.

Subject matter expert 1has over twenty years of experience working as a trainer of school psychology in Utah and has fifteen publications on data use. Subject matter expert 2 has over 10 years of experience as a professor in Pennsylvania, received a lifetime achievement award in educational psychology, and has over 20 publications in school psychology. The third subject matter expert has over 10 years of experience working as a professional psychologist in Kansas and has over 30 publications in educational psychology.

Measures

Demographic questionnaire (Appendix B): The demographic survey asked teachers to identify their gender, ethnicity, highest degree received, years of experience, data-related courses, and specialization (e.g., general or special education).

Comfort with Data Use Scale (Appendix F; Doll et al., 2005): The Comfort with Data Use Scale is a modified four-item scale of teacher's self-reported comfort of data use. This measure's items are scored on a scale from 0 to 10 with a zero representing an item that "Does not describe me at all" and a 10 representing "Describes me accurately." The sum across the four items ranged from 0 to 4.

NU Data Knowledge Scale (See Appendix C; Sikorski et al., 2014): The NU Data Knowledge Scale is a 30-item, four choice, multiple-choice instrument assessing teachers' knowledge of the six databasics. Each item is scored as binary with (1) for correct responses and (0) for incorrect responses. Each item has only one correct response. The sum across all items will range from 0 to 3.

DataBasics Matrix (See Appendix D): A matrix was constructed with a row for each of the NU Data Knowledge Scale's 30-items and a column for each of the six databasics. The subject matter experts used this matrix to code the NU Data Knowledge Scale items by which of the six databasics they reflect. A single scale item could be categorized in more than one databasic due to the high inter-correlation between the databasics. The completed matrices provided evidence of the content validity of the NU Data Knowledge Scale by incorporating expert opinion into the measure's items.

Procedures

This study received approval from the Institutional Review Board to recruit teachers and subject matter experts from surrounding Nebraska schools and to conduct the proposed analyses described below.

The subject matter experts were instructed to categorize each item of the NU Data Knowledge Scale on the DataBasics Matrix according to the DataBasics they were best categorized by. The principal investigator conducted brief analyses of the subject matter expert ratings and created rubric to discuss items that was color coated to assist discussion.

Then the subject matter experts met with the principal investigator for a two-hour adobe connect meeting and conference call to provide their expert opinion of measure's items, and to discuss how well Knowledge Scale items fit the databasics conceptual framework. Overall, the subject matter experts had total agreement ("green") on 17 test items, partial agreement ("yellow") on 10 items, and total disagreement ("red") on three items. The discrepancies in matrix coding were discussed during the adobe connect meeting starting with the red items, then the yellow items, and finally the green items. Every item was discussed and disagreements were reconciled during the two hour meeting. During this meeting, it was decided that Item 6 needed to be totally revised due to its overlap with other test items. This item was rewritten by the subject matter experts, and the revised measure was sent into IRB for approval with other minor test revisions. The subject matter experts were given an honorarium of $250 dollars for their time.

The NU Data Knowledge Scale was disbursed to 215 teachers after the subject matter experts' recommendations and revisions were incorporated into the measure and

approved by the IRB. The Qualtrics survey included a digital consent form, demographics survey, Comfort with Data Use survey, and NU Data Knowledge Scale. Consenting teachers were entered into a drawing for one of five $50 amazon gift cards that was conducted after all data were collected. The 215 teacher participants were recruited over a one week period in which the survey was open.

Data Analysis

Data in this study were teachers' responses to the 30 NU Data Knowledge Scale items, 14 predictor variables were used to predict teacher performance on the knowledge scale which included: gender, years of experience, size of classroom, highest degree received, Comfort with Data Use, and teaching assignment(s). Data were recorded electronically and organized from the Qualtrics file. Teacher responses on the knowledge scale were scored dichotomously (1 for correct responses, 0 for incorrect responses) for analysis of the coefficient alpha, item-total correlations, corrected point-biserial correlations, and weighted least squares means of variance.

Traditional item statistics analyses were used to address the first research goal about the psychometric properties of the scale (Thorndike & Thorndike-Christ, 2010). Descriptive statistics included the mean, standard deviation, range, kurtosis, skewness, and homogeneity of scores (see Table 4:1). These descriptive statistics were used to identify potential outliers within the data set. Kuder-Richardson 20 analysis was used to assess the overall internal consistency of teacher's scores on the instrument. KR-20 reflects the internal consistency of the teachers' scores on a binary-scored measure, and is an estimate of the measure's reliability. Acceptable KR-20 scores vary by application. For example, a KR-20 of .7 is acceptable when a measure is being developed or refined,

but not acceptable for high stakes applications such as testing intelligence. Since the NU Data Knowledge Scale is still in the development phase, a KR-20 score of .7 or above is expected for the measure to show evidence of internal consistency (Thorndike & Thorndike-Christ, 2010).

To assess the capacity of items to discriminate teachers' knowledge of the databasics, the item-total correlations of the items were calculated. If the items of the NU Data Knowledge Scale are discriminative, these would have item-total correlations above .2. At the item level, corrected point-biserial correlations and *p*-values were calculated to determine the discriminative properties of the measure at the item level. Corrected point-biserial correlations determine how well test items and distractors discriminate teachers' knowledge (Keppel & Wickens, 2004). Evidence that the NU Data Knowledge Scale discriminated teachers' knowledge, would require that the corrected point-biserial correlations for the correct answer to be correlated with the overall score and above .2, and distractors be negatively correlated with the overall score. *P*-values were used to determine the traditional difficulty of items by showing the proportion of teachers who correctly completed an item. Optimal *p*-values for the test items would range between .15 and .85 with an average between .4 and .6 (Thorndike & Thorndike-Christ, 2010). *P*-values range from 0 to 1 with scores closer to 1 reflecting easier items.

Subject matter expert agreement and WLMSV analyses were examined to address the second goal determining how well the exploratory factor analysis matches the databasics theoretical framework for items measuring data literacy. To determine whether subject matter experts' ratings aligned with the test blueprint of the knowledge scale, percentage of agreement and un-weighted kappa scores of agreement were calculated.

The magnitude of agreement between observers and the test blueprint was calculated for each databasic. Prior to consensus building, it was expected that kappa scores of agreement among the three raters would be moderate (i.e., kappa greater than .41; percentage of agreement above .80).

The dimensionality of the scale was determined through the weighted least mean squares of variance analysis and the computation of parallel analysis. An oblimin rotation with a zero delta weight was used to account for correlations between items and the databasics (Tabachnick & Fiddell, 2001). A direct oblimin rotation allowed for stronger eigenvalues between items due to the expected correlation between databasics. A tetrachoric coefficient was used to determine the correlation between items since they were scored dichotomously and assume that the data were normally distributed (Sun et al., 2007). As a result, a WLSMV analysis was conducted to determine the factor structure because it is the most effective for analyzing categorical data (Brown, 2006). Parallel analysis was used to determine cutoff scores for eigenvalues for the sample data. Parallel analysis provides more stringent eigenvalue cutoff scores than Kaiser's (1960) criteria when determining which factors to maintain and is more objective than visually interpreting the scree plot (Hayton et al., 2004).

A parallel analysis was conducted by first generating random data that matched the number of observations (n) and variables (e.g., 30-items) to create a random data set outlined by Hayton, Allen, and Scarpello (2004). Then the eigenvalues were extracted by noting the results of principle component analyses of 5000 randomly generated data sets. The mean of the 95th percentile of all observed eigenvalues were used to create a vector of average 95th percentile eigenvalues in equal size to the number of variables and

diminishing in value. Finally, these eigenvalues were plotted on top the real data set. Factors were retained on the real data set only if they were greater than the eigenvalues of the randomly generated data.

Scree plots with the eigenvalues from the original data and parallel analyses were examined to determine how many factors should be retained. Examination of the scree plot and eigenvalues were then tested to determine if they were significant. The residual variance of each item was examined to determine what percentage of each item was accounted for by the latent variable data literacy. The NU Data Knowledge Scale factor structure was analyzed to see if it had a non-significant Chi-Square test, goodness of fit above .90, comparative fit index greater than .90, a Tucker-Lewis Index above .95, and a standardized root mean square below .05 (Steiger & Lind, 1980).

To determine the best model for describing the relation between years of experience, gender, class size, education level, Comfort with Data Use Scale, and teacher assignment with teachers' performance on the NU Data Knowledge Scale (Research Goal 3) all possible regression analyses were conducted. Mallow's Cp (Mallows, 1973) was used to determine the best subset of the 14 predictors for predicting teachers' scores on the NU Data Knowledge Scale. Mallow's Cp is a measure that accounts for differences in the residual term by adding more predictors and observing the changes in magnitude (Lieberman & Morris, 2008). The Mallow's Cp statistic was analyzed by comparing the number of predictors plus one (p+1) criterion for determining which model best fit the data. Models with a Mallow's Cp score closest to the p+1 and highest adjusted R^2 criterion were determined to have the best model fit.

CHAPTER FOUR: RESULTS

The following chapter describes the results of this study. First, preliminary analyses will examine the agreement, model fit, and significance of results will be discussed. Then, the results of each research goal will be described.

Preliminary Analyses

Assumptions for Data Analysis. Data were examined to ensure that the statistical assumptions were met for continued analyses. There were no missing data in any of the 215 teacher cases who successfully completed the survey. Because items on the NU Data Knowledge Scale were scored dichotomously (1 for correct, 0 for incorrect) and had a natural ordering, they were viewed as an ordinal variable. A tetrachoric correlation was used during the factor structure of the measure which assumed that the underlying (latent) responses were normally distributed. Some demographic data were collapsed due to low representation in the sample. For instance, Speech/Language specialists were combined with special education teachers; Associates and Bachelors level teachers; and EdS and PhD. teachers were combined during the all possible regressions analyses. Years of experience and size of classroom were viewed as ordinal variables; while gender, classroom assignment, and highest degree received were analyzed as nominal variables. The Comfort with Data Use measure was analyzed as an interval variable.

Research Goal 1

The Kuder-Richardson 20 statistic was used to estimate the internal consistency of the NU Data Knowledge Scale. The ltm package in R was used to determine the coefficient alpha of the NU Data Knowledge Scale and is displayed in Table 4:1.

Table 4:1.

Kuder-Richardson 20 scores for participants

KR-20	
Raw KR-20	Standardized KR-20
.6135	.66279

The observed KR-20 score of .6135 was below the standard value for developing measures .7. The overall KR-20 score is within the "acceptable" range for a measure, but caution would be advised for using the NU Data Knowledge Scale for making high stakes decisions about teachers' knowledge of the NU Data Basics (Thorndike & Thorndike-Christ, 2010).

The corrected item-total correlations for each item were analyzed using the baseR package in R-Studio. The corrected item-total correlations were calculated to determine how well individual test items correlate with the overall Knowledge Scale score and are displayed in Table 4:2.

Table 4:2.

Corrected Item Total Correlations and If Item Deleted Analyses

Item	Mean if Item Deleted	Variation if Item Deleted	Correlation with Total	Alpha if Item Deleted
When is an intervention evidence-based	19.9256	12.5832	.1386	.6087
What is baseline data?	19.8279	12.5637	.2691	.6018
Which of the following is an example…?	19.8279	12.4889	.3142	.5992

Which is the following is an observable…?	19.8000	12.8617	.1436	.6098
How would you translate a student's…?	20.1674	12.1588	.1976	.6030
Which of the following best describes…?	20.1256	12.2599	.1740	.6057
Which of the following…example in math…?	19.8698	12.5063	.2224	.6027
In addition to class grades, how could you…?	19.8698	12.3942	.2759	.5988
What should you do before collecting…?	20.1209	12.1348	.2133	.6011
What is an important consideration…?	20.3535	13.3511	-.1437	.6418*
You have been monitoring the number…?	20.3349	12.2705	.1608	.6074
Pat is constantly disrupting class by… ?	20.5535	12.6315	.0945	.6133
Your team has collected data on a…?	20.1954	12.2981	.1531	.6083
A teaching team is worried about a student…?	19.9861	12.2101	.2400	.5989
A third grade teacher surveyed…?	20.0140	12.2755	.2030	.6025
A sixth grade teacher surveyed… ?	20.1535	12.0090	.2452	.5973
After collecting data on a student's…?	20.0837	12.1145	.2295	.5994
A special education team met to… ?	20.3023	12.6512	.0493	.6205*
How could you describe…?	19.9767	12.3499	.1955	.6034
A student with cognitive disabilities…?	19.8512	12.1927	.4161	.5903
Your team was referred a student who…?	20.0605	11.8234	.3331	.5878
A teacher has been implementing an…?	20.3070	12.4287	.1130	.6131
You have collected 5 days of …?	20.2977	12.5091	.0896	.6159*
A team of teachers met briefly…?	19.8791	12.4339	.2439	.6008
You have been collecting data on…?	20.1023	11.9147	.2872	.5926
A middle school team has been … ?	20.1395	12.7842	.0166	.6236*

How would you describe the trend…?	19.8140	12.4886	.3642	.5981
An Art teacher wanted to record…?	19.9581	12.2179	.2556	.5978
Midway through the year…?	20.2465	12.4203	.1149	.6129
You are working with a first year…?	20.1116	12.1370	.2158	.6010

** Indicates KR-20 score would improve if item was deleted.*

Corrected item-total correlations ranged from -.1437 to .4161 for individual test items. All but Test Item 10 were positively correlated with the overall test score. This indicates that teachers who answered Item 10 correctly did worse on the overall measure, on average, than teachers who answered the item incorrectly. Removing Item 10 from the analyses would increase the alpha score by .0283. Hence, Item 10 should be revised or removed from future iterations of the measure. It is important to note that Item 10 addressed DataBasic 6, and the revised Item 10 should address this databasic as well.

Sixteen items (e.g., items 2, 3, 7, 8, 9, 14, 15, 16, 17, 20, 21, 24, 25, 27, 28, and 30) had corrected item-total correlations above the expected value of .2. These items correlated moderately with the overall test score and contributed to overall consistency of the measure. Items with an item-total correlation below .2 and above 0 were not very discriminative of teachers' overall knowledge of the databasics and should be revised in future iterations of the NU Data Knowledge Scale.

Additional analyses at the item level revealed problems with some of the distractors for items on the NU Data Knowledge Scale. The corrected point-biserial correlations of the distractors and correct answers of the Knowledge Scale items revealed that almost half of the items had one or more distractors that were positively correlated

with the overall score. These corrected point-biserial correlations can be observed in Table 4:3.

Corrected Point-Biserial Correlations for Knowledge Scale Items

	Corrected Point-Biserial Correlations by Response				
Item	A	B	C	D	*p*-value
When is an intervention evidence-based	-.0940	.0114	-.0712	**.0785**	.8458
What is baseline data?	-.1240	-.0705	**.2152**	-.0742	.9393
Which of the following is an example…?	-.0121	**.2802**	-.1561	-.1093	.9346
Which is the following is an observable…?	**.2360**	N/A	-.3090	-.0482	.9533
How would you translate a student's…?	-.0498	-.1313	**.1233**	-.0535	.5935
Which of the following best describes…?	.0251	**.1856**	-.1365	-.1052	.6262
Which of the following…example in math…?	-.0596	-.0628	-.1000	**.3285**	.8738
In addition to class grades, how could you…?	-.0019	**.3687**	-.0751	-.1760	.8692
What should you do before collecting…?	-.0770	-.3238	**.2324**	-.0404	.6589
What is an important consideration…?	-.0690	-.0174	-.1507	**.1125**	.3879
You have been monitoring the number…?	-.1490	-.2267	**.1930**	.1127	.4019
Pat is constantly disrupting class by… ?	.1378	**.1181**	-.3984	-.0070	.2383
Your team has collected data on a…?	.0260	-.1578	**.0475**	.0312	.5467
A teaching team is worried about a student…?	-.2420	.0301	**.3574**	-.3376	.7477
A third grade teacher surveyed…?	.0341	-.1691	**.1155**	-.1712	.7570
A sixth grade teacher surveyed… ?	-.3542	**.3647**	.0645	-.3866	.5701
After collecting data on a student's…?	-.2347	**.3698**	-.3934	.0569	.6449
A special education team met to… ?	-.0434	-.0796	**.0982**	.0405	.4346

How could you describe…?	.0241	-.4387	**.4018**	-.1146	.7196
A student with cognitive disabilities…?	-.5805	**.6162**	-.1100	-.1539	.8411
Your team was referred a student who…?	-.6011	-.0558	**.4956**	-.0791	.6542
A teacher has been implementing an…?	.0055	**.2243**	-.2966	-.1623	.4252
You have collected 5 days of …?	-.1048	**.1345**	.0160	-.1068	.4252
A team of teachers met briefly…?	**.1839**	-.1826	-.1925	-.0395	.7944
You have been collecting data on…?	-.0456	-.0626	-.2113	**.2152**	.6215
A middle school team has been … ?	.0279	-.1604	**.2540**	-.2425	.6262
How would you describe the trend…?	-.1542	-.5609	**.6285**	-.1912	.8411
An Art teacher wanted to record…?	-.4819	**.4712**	-.1615	-.0965	.7056
Midway through the year…?	-.1811	.0870	**.1795**	-.2409	.4860
You are working with a first year…?	**.1277**	-.1200	-.1730	-.0056	.5607

For sixteen items (e.g., items 2, 3, 4, 5, 7, 8, 9, 10, 20, 21, 24, 25, 26, 27, 28, and 30) only the correct answer was positively correlated with the overall score. Of these sixteen items, only ten of them had a corrected point biserial correlations above .2. This is indicative of the NU Data Knowledge Scale items having multiple distractors that were perceived as being correct by teachers. Item 4's distractor "b" was score as an N/A since no one selected it and should be revised in future revisions of the measure.

The proportion of teachers answering each item correctly was analyzed to determine the discriminative properties of the instrument. The proportion scores for the items ranged from .2150 to .9674 with an average of .6922. This means that, averaging across items teachers, answered correctly 69.22 percent of the time. The overall average

of the *p*-values scores is higher than the expected .4 to .6 range, and the maximum proportion of an item (.9694) was higher than the expected .85. Items that scored above the .85 range are not very discriminative since a majority of participants answered the item correctly regardless of their overall ability. Eight items scored above the .85 range, including items 2, 3 4 7, 8, 20, 24, and 27. These items may be considered too easy for participants and do not discriminate teachers' knowledge of the databasics. These eight items do not discriminate by participants' by their data literacy, but may increase lower achieving participants' confidence while they complete the measure. The correlation of distractors may be higher in untrained teachers than in teachers who were trained in the nuances of data literacy.

The descriptive statistics for the NU Data Knowledge Scale and Comfort with Data Use were obtained by using the psych package in R for participants who completed the entire survey (e.g., 215 out of 287 teachers). A total of 215 teacher scores were analyzed to obtain the mean, standard deviation, range, skewness, and kurtosis of the overall scores. Descriptive statistics for all measures' means, standard deviations, and ranges are included in Table 4:4. The NU Data Knowledge Scale, Comfort with Data Use, and demographic questionnaire were entirely completed by 215 certified teachers (38 male, 177 female). Seventy-two cases were omitted from the analyses when teachers failed to complete the survey, resulting in a completion percentage of 74.92 percent. The mean score on the NU Data Knowledge Scale was 20.77 with a standard deviation of 3.62, suggesting that teachers' knowledge of the databasics varied on the 30-item measure. The mean score on the Comfort with Data Use scale was 27.87 out of 40 with a

standard deviation of 9.02 suggesting teachers expressed moderate comfort with their ability to use data. The results of the analyses are displayed in Table 4:4.

Table 4:4

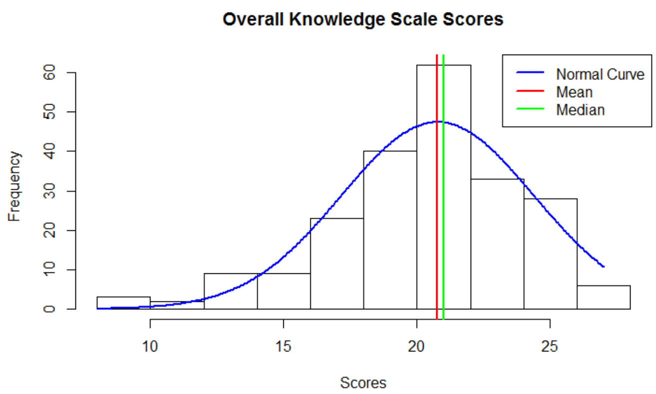
Descriptive Statistics for Overall Teacher Scores

Measure	N	Mean	Sd	Min	Max	Range	Skewness	Kurtosis
NU Data Knowledge Scale	215	20.77	3.62	8	27	19	-.81	.91
Comfort with Data Use	215	28.78	7.21	1	40	39	-.79	.64

The overall mean on the NU Data Knowledge Scale was (20.77) and standard deviation (3.62) of the measure allows for growth in participant scores across administrations on the 30-item measure. The ceiling of the test is two standard deviations above the mean which is desirable for an instrument that could be administered multiple times. There was also adequate range in the overall scores. Out of the 215 participants, no one received a perfect score of 30 or a totally incorrect score of 0. The minimum score of 8 is above the probability of random guessing. The data were negatively skewed, however, in that more participants scored above the mean. The score distribution compared to a normal curve can be seen in Figure 4:1.

Figure 4:1

Histogram and normal curve of overall scores on NU Data Knowledge Scale



The distribution of scores appear to be normally distributed with a few participant scores negatively skewing the overall data. The scores had a kurtosis value .91 and skewness value of -.81 which are both between -1 and +1 showing that the data are normally distributed (Thorndike & Thorndike-Christ, 2010). As a result, the distribution of the data should not negatively impact interpretation of other analyses.

Scores on the Comfort with Data Use Scale ranged from 0 to 40 with an mean of 28.78 and standard deviation of 7.21. There was wide variation in teachers' self-ratings of their comfort with data use. Comfort with data used scores were negatively skewed and not normally distributed.

Research Goal 2

The alignment of the NU Data Knowledge Scale items with the databasics was determined by analyzing the categorizations of three subject matter experts. The percentage of agreement and unweighted kappa scores were used to determine the inter-rater agreement and are displayed in Table 4:5.

Table 4:5

Percentage of agreement and kappa scores of the subject matter experts

	SME 2		SME 3	
	% Agreement	Kappa Statistic	% Agreement	Kappa Statistic
SME 1	81.7	.421*	80	.439*
SME 2			79.4	.455*

* Denotes significance (*p* < .05)

Overall, the agreement between subject matter experts was good with scores ranging from 79.4 to 81.7 percent and kappa scores that were acceptable ranging from .421 to .455. These scores are indicative of moderate levels of agreement between the raters' categorizations of the measures' items by databasics and were within the expected range for this study.

After consensus was established between raters, the SME revised blueprint (see Appendix E) was compared to the original test blueprint developed prior to the study and is displayed in Table 4:6.

Table 4:6

Overall agreement and kappa scores between aggregate SME scores and test blueprint

	% Agreement	Kappa
Overall	70.6	.439*

* Denotes significance (*p* < .05)

The overall agreement between the SME revised blueprint and original test blueprint of 70.6 percent and a kappa of .439 are acceptable. These scores reflect moderate levels of agreement between the subject matter experts and the original test blueprint. Test items on the NU Data Knowledge Scale were categorized with a high

level of agreement by databasics which is indicative of items being clearly defined to assess the databasics.

The revisions to the test blueprint resulted in two databasics being underrepresented in the Knowledge Scale. These results are displayed in Table 4:7.

Table 4:7

Revised test blueprint

	Scale Items	Percentage
DataBasic 1	4, 8, 9, 29	13.33
DataBasic 2	6, 8, 9, 12, 16, 20, 21, 24, 28, 29	33.33
DataBasic 3	5, 11, 13, 15, 17, 23	20.00
DataBasic 4	11, 19, 20, 22, 23, 27	20.00
DataBasic 5	2, 3, 14, 18, 19, 24, 25, 26, 30	30.00
DataBasic 6	1, 7, 10, 26, 30	16.67

As displayed in Table 4:7, DataBasics 1 and 6 are underrepresented in the NU Data Knowledge, while DataBasics 2 and 5 are overrepresented in the SME revised blueprint. Ideally, each databasic would be represented in a minimum of 20 percent of test items (e.g., 6 out of 30).

The dimensionality of the 30-item measure was analyzed using a limited-information weighted least squares mean and variance (WLSMV) adjusted estimation analysis in m*PLUS*. A direct oblimin rotation was used to determine the factor structure of the measure. When conducting these analyses, the resulting factor loadings' significance is determined by the *p*-value which differs from principal axis factoring

criteria of loadings greater than .32. Examination of the factor loadings and results from the parallel analysis were used to determine the dimensionality of the scale and are displayed in Table 4:8 and Figure 4:2.

Table 4:8

Standardized Single Factor Loadings on Probit Scale

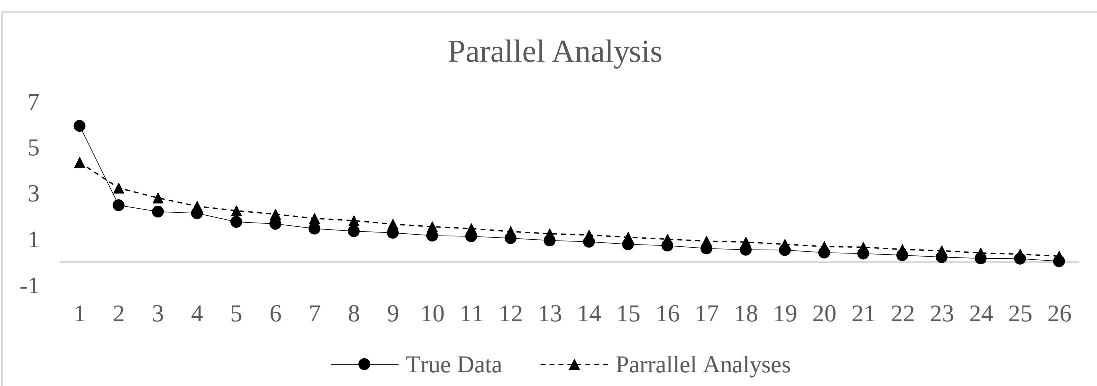
	Factor Loadings
Item	Data Literacy
When is an intervention evidence-based	.266*
What is baseline data?	.560*
Which of the following is an example…?	.799*
Which is the following is an observable…?	.414*
How would you translate a student's…?	.268*
Which of the following best describes…?	.244*
Which of the following…example in math…?	.466*
In addition to class grades, how could you…?	.542*
What should you do before collecting…?	.372*
What is an important consideration…?	-.222*
You have been monitoring the number…?	.248*
Pat is constantly disrupting class by… ?	.315*
Your team has collected data on a…?	.261*
A teaching team is worried about a student…?	.482*
A third grade teacher surveyed…?	.349*
A sixth grade teacher surveyed… ?	.416*

After collecting data on a student's…?	.384*
A special education team met to… ?	.065
How could you describe…?	.322*
A student with cognitive disabilities…?	.757*
Your team was referred a student who…?	.568*
A teacher has been implementing an…?	.228*
You have collected 5 days of …?	.211*
A team of teachers met briefly…?	.429*
You have been collecting data on…?	.473*
A middle school team has been … ?	.095
How would you describe the trend…?	.901*
An Art teacher wanted to record…?	.497*
Midway through the year…?	.234*
You are working with a first year…?	.357*

* Denotes significance at the p = .05 level

Figure 4:2

Scree plot of parallel analysis



The scree test of the parallel analysis revealed that a single factor solution fit the data best with one sample-data eigenvalue greater than the mean of the 95th percentile of eigenvalues from the parallel analysis. Standardized factor loadings for a single factor solution ranged from -.222 for item 10 to .950 for item 26. Overall, 28 items were determined to fit the single factor solution well. All but two items (18 and 35) had p-values less than .05 indicating that they were significantly different from 0 (i.e., the null model). The single factor solution had a non-significant Chi-Square test (p = .1263) which indicates that the sample data are not significantly different from the single factor model. The root mean square error of approximation (RMSEA) was significant and below .05 (RMSEA = .019, p > .999, CI .000 - .032) indicating failure to reject the null hypothesis of approximate fit. The Tucker-Lewis index (TLI) and comparative fit index (CFI) were both below the expected value of .90 (.870 and .879, respectfully) which does not support the single factor solution. Both TLI and CFI are recommended to be above .90 to indicate strong model fit and that the observed sample is different from an independent sample (Hu & Bentler, 1999).

The proportion of variance in items accounted for by the single factor solution can also be calculated. The R-squared for each item and the single factor solution are displayed in Table 4:9.

Table 4:9

R-Square scores for items in single factor solution

Observed Variable	R^2	S.E.	Est./S.E.	Two-Tailed P-Value	Residual Variance
When is an intervention evidence-based?	.071	.068	1.039	.299	.929

What is baseline data?	.314	.159	1.971	.049*	.686
Which of the following is an example…?	.607	.187	3.254	.001*	.393
Which is the following is an observable…?	.171	.118	1.449	.147	.829
How would you translate a student's…?	.072	.054	1.332	.183	.928
Which of the following best describes…?	.059	.048	1.226	.220	.941
Which of the following…example in math…?	.218	.123	1.773	.076	.782
In addition to class grades, how could you…?	.293	.145	2.022	.043*	.707
What should you do before collecting…?	.139	.068	2.036	.042*	.861
What is an important consideration…?	.052	.044	1.187	.235	.948
You have been monitoring the number…?	.062	.046	1.350	.177	.938
Pat is constantly disrupting class by… ?	.099	.051	1.953	.051	.901
Your team has collected data on a…?	.068	.051	1.334	.182	.932
A teaching team is worried about a student…?	.232	.096	2.411	.016*	.768
A third grade teacher surveyed…?	.122	.073	1.658	.097	.878
A sixth grade teacher surveyed… ?	.173	.075	2.298	.022*	.827
After collecting data on a student's…?	.147	.075	1.971	.049*	.853
A special education team met to… ?	.004	.014	.314	.754	.996
How could you describe…?	.110	.073	1.503	.133	.890
A student with cognitive disabilities…?	.573	.159	3.609	.000*	.427
Your team was referred a student who…?	.323	.112	2.885	.004*	.677
A teacher has been implementing an…?	.052	.044	1.171	.242	.948
You have collected 5 days of …?	.044	.041	1.094	.274	.956
A team of teachers met briefly…?	.184	.114	1.610	.107	.816

You have been collecting data on…?	.224	.086	2.619	.009*	.776
A middle school team has been … ?	.009	.020	.459	.646	.991
How would you describe the trend…?	.812	.191	4.258	.000*	.188
An Art teacher wanted to record…?	.247	.096	2.558	.011*	.753
Midway through the year…?	.055	.043	1.278	.201	.945
You are working with a first year…?	.127	.073	1.749	.080	.873

* Denotes significance ($p < .05$)

Overall, only twelve items had a significant amount of their variance explained by the single latent factor. The R^2 ranged from .004 to .812 with an average R^2 of .1887. This is not indicative of the single factor solution accounting for a large percentage of variance within the overall scores.

Research Goal 3

In order to determine which demographic and background characteristics best predicted scores on the NU Data Knowledge Scale, a series of 1024 models (all possible models) were run using R to compare the regressions from teacher years of experience; number of students in their classroom; teacher self-ratings of knowledge, data use, ability to graph, and comfort sharing data; gender; highest degree received; teaching assignment; and grade level. The top five models and their predictors are displayed in Table 4:1. The number of predictors (p) are compared to Mallow's Cp.

Table 4:10

All possible regressions table sorted by Difference in Mallow's Cp and (p+1)

Model	R2	AdjR2	Cp	Predictors (p)	Difference between (p+1) and Cp	Omnibus Predictors
1024	.1736	.1202	14	13	0	ABCEFGHIJ
1020	.1731	.1240	13.1251	12	.1251	ACDEFGHI
1022	.1722	.1231	13.3368	12	.3368	ABCDEGHIJ
1018	.1720	.1228	13.4005	12	.4005	ABCDEGHIJ
1003	.1718	.1269	12.4469	11	.4469	ACEFGHIJ

#A=Years of experience
#B=Students
#C=Know a variety of ways to collect
#D=Use data daily
#E=I am able to graph
#F=Comfortable sharing data
#G=Female (0 = Male, 1 = Female [ref])
#H=Deg (1 = Bachelor's or less [ref], 2 = Master's, 3 = More)
#I=Assignment (1 = Gen Ed [ref], 2 = SPED, 3 = Other)
#J=Grade level (1 = Elem [ref], 2 = Secondary, 3 = Other)

These top models were selected by comparing the Mallow's Cp statistic with the p+1 criterion. The top model is selected when the Mallow's Cp statistic is close to the number of predictors plus one. A model with a Cp greater than p+1 is said to be overspecified while a Cp less than p+1 is underspecified (i.e., an important predictor has been omitted). Despite a large number of predictors, the top model [1020] only accounted for 17.31 percent of the teachers' score on the knowledge scale. Note model 1024 is the most saturated model (i.e., contains all of the predictors) and is excluded from being interpreted using Mallow's Cp because the difference between the number of predictors and (p+1) is always zero. Despite the model containing 12 predictors, the next best model (1020) has a low R^2. Despite the model being significant ($F = 3.379$, $p < .001$), the resulting R^2 lacks practical significance. The predictors in this model and their R^2 can be

viewed in Table 8. This model is not indicative of the predictors having much influence over teachers' knowledge of the databasics. Additionally, the shrinkage in the adjusted R^2 to .1240 indicates that the model is over fitting the sample and may lack generalizability to the larger teacher population.

Table 4:11

Omnibus R^2 for predictors in Model 1020

Model		df	R-square	*p*-value
1020		12	.1718	.0001
	Years of experience	1	.0087	.0910
	Number of students	1	.0010	.5726
	Know a variety of ways	1	.0032	.3047
	I am able to graph	1	.0205	.0096
	Comfort with sharing data	1	.0010	.5680
	Gender (Male)	1	.0086	.0916
	Degree (Masters, EdS & PhD)	2	.02370	.0207
	Teaching Assignment (Gen., Other)	2	.0027	.6348
	Grade Level (Secondary, Other)	2	.0117	.1141

There is ample evidence that model 1020 though significant, is not generalizable to the broader teaching population. Many predictors included in model 1020 were not significant in predicting teachers' overall scores. As a result, it can be inferred that years of experience, gender, teaching specialization, highest degree received, and teachers'

self-ratings on the Comfort with Data Use scale do not predict teachers' knowledge of the databasics or their data literacy in a meaningful way.

An alternative method for selecting the best fitting model in all possible regressions is to examine the adjusted R^2 in addition to reviewing the Mallow's Cp. In this scenario, model 923 is the best fitting with a Cp less than one and having a relatively high Adjusted R^2 of .1296. These results are displayed in Table 4:12.

Table 4:12

Omnibus R^2 for predictors in Mode 923

Model		df	R^2	*p*-value
923		10	.1703	.0002
	Years of Experience	1	.0061	.0954
	Know a variety of ways…	1	.0017	.3861
	I am able to graph	1	.0149	.0098
	Gender (male)	1	.0007	.8189
	Degree (Masters, EdS & PhD)	2	.0173	.0206
	Teaching Assignment (Gen., Other)	2	.0015	.7086
	Grade level (Secondary, other)	2	.0076	.1782

Overall, model 923 had a similar R^2 and slightly higher Adjusted R^2 than the most saturated model 1024 and model 1020 and is significant ($F = 4.186$, $p < .0001$). As displayed in Table 4:12, many individual predictors in model 923 lacked significance. However, with ten predictors, model 923 is still at risk for overfitting the data. A comparison of the normal plots and residual plots of the data for models 1020 (see Appendix H) and model 923 (see Appendix I) show that both models are not curvilinear.

In conclusion, model 923 and 1020 are significant, however, they are not very meaningful when generalized to the larger teacher population due to their low adjusted R^2. The predictors in this study did not discriminate teachers' scores on the NU Data Knowledge Scale in a meaningful way.

Diagnostics for Model 923. Model 923 was chosen as the best fitting model because of its statistical characteristics. An examination of the studentized residual vs fitted plot confirms that values are randomly distributed about the line of the mean studentized residuals which rejects the presence of heteroscedasticity. Additionally, the normality of residuals were visually checked and determined to be within normal limits (see Appendix I; Tabachnich & Fidell, 2013).

A closer examination of the Residual vs. Leverage plot (see Appendix I), reveals three participants' (61, 106, and 131) scores negatively affect (leveraged) the regression line. These outliers were further examined by comparing their Cook's Distance values (see Appendix J), and appeared significantly out of line with other subjects. The dfBetas were calculated for these subjects, and cutoff score for influence was calculated using Neter, Wasserman, and Kutner's (1989) criteria for *adjusted cutoff scores.* The adjusted cutoff score of $2/\sqrt{N}$ (.1363) was used to determine if these outliers influenced the intercept and the regression coefficient of Model 923. The dfBetas for Participants 61, 106, and 131's influence on Model 923's predictors and intercept are displayed in Table 4:13. As can be seen, the intercept dfbetas for Participants 106 (dfBeta = -0.5563) and 131 (dfBeta = -0.2322) were above the adjusted cutoff criteria. With respect to the predictors, for Participant 106 the dbBetas for the predictors Years, Variety, Graph, and Gender exceeded the adjusted criterion, whereas for Participant 131 the dfBetas for

Years, Gender, and Teaching Assign all exceeded the criterion. Participant 61 had only two predictors whose dfBetas exceeded the criterion (i.e., Variety, Graph) although a third predictor, Gender, was very close. Nevertheless, it was decided to retain these cases in model 923 due to their marginal influence on the regression coefficient.

Table 4:13

DfBeta scores for participants

Participant	Intercept	Years	Variety	Graph	Gender	Degree	Teaching assign	Grade level
61	.1244	.0012	.1569*	.4312*	-.1323	-.0925	-.0351	-.0849
106	-.5563*	-.3924*	.8749*	-.7336*	.6155*	.0408	-.04347	.1147
131	-.2322*	.2161*	-.1298	.06334	.4272*	.0928	-.1966*	-.1282

*above cutoff criteria of $2/\sqrt{N}$ (.1363)

The collinearity of Model 923 was examined by calculating a correlation matrix and computing the tolerance and variance inflation factor (VIF). A correlation matrix (see Table 4:14) was also created to examine the correlation between independent variables. Table 4:14 displays the correlation coefficient for each predictor variable above the diagonal and the p-value for the correlation coefficients below the diagonal.

Table 4:14

Model 923 Correlation Matrix

	A	B	C	D	E	F	G	H	I	J
A	*****	0.044	0.18	0.035	0.052	0.096	-0.02	0.495	0.031	0.053
B	0.52	*****	0	-0.086	0.002	-0.038	-0.007	0.073	-0.155	0.091
C	0.008*	0.997	*****	0.675	0.567	0.724	0.043	0.231	0.094	-0.107
D	0.614	0.208	<0.001*	*****	0.584	0.652	0.201	0.16	0.146	-0.212
E	0.448	0.973	<0.001*	<0.001*	*****	0.624	0.112	0.173	-0.003	-0.129
F	0.159	0.584	<0.001*	<0.001*	<0.001*	*****	0.067	0.168	0.078	-0.145

G	0.766	0.916	0.526	0.003*	0.102	0.325	*****	0.027	0.073	-0.21
H	<0.001*	0.285	0.001*	0.019*	0.011*	0.014*	0.693	*****	-0.033	0.033
I	0.647	0.023*	0.168	0.032*	0.971	0.252	0.284	0.629	*****	0.382
J	0.441	0.185	0.116	0.002*	0.058	0.034*	0.002*	0.632	<0.001*	*****

Table 4:14 revealed several instances in which variables were significantly correlated with each other. However, this is to be expected as C (Use Data Effectively), D (Comfort with Data Use), E (I can graph…), and F (Comfortable sharing data…) are from the same measure and were used to determine teachers' self-appraisals of their data-use familiarity and skills. Years of expereince was significantly correlated with Knowing a Variety of Ways to Collect Data and type of degrewe, and classroom size was corrleated with teaching assignment. To determine if these linear relations impact the regression coffecient, the tolerance and VIF values were examined. The results of these analyses are included in Table 4:15. The assumptions of non-multicollinearity were met as evidenced by VIF values less than 10 and tolerance values greater than 0.2 (Menard, 1995; Myers, 1990).

Table 4:15

Tolerance and Variance Inflation Factor for Mode 923

Model 923 Predictor Tolerance and VIF Table							
	B	SE B	Beta	Tol.	VIF	T	Sig T
Years of Experience	-0.338	0.2286	-0.101	0.8535	1.1716	-1.4790	0.1407
Know a variety of ways…	0.1877	0.1714	0.0879	0.7842	1.2752	1.0950	0.2748

I am able to graph	0.288	0.1135	0.1993	0.7933	1.2606	2.5370	0.0119
Gender (male)	1.2013	0.63	0.127	0.9329	1.0719	1.9070	0.0579
Degree (Masters, EdS & PhD)	1.0693	0.3966	0.2027	0.9090	1.1001	2.6960	0.0076
Teaching Assignment (Gen., Other)	0.3982	0.3487	0.81334	0.9291	1.0763	1.1420	0.2549
Grade level (Secondary, other)	-0.3367	0.3175	-0.0771	0.9066	1.103	-1.0600	0.2902

Overall, Model 923 meets criteria for model fit. The residuals are normally distributed, outlier data do not influence or leverage the regression coefficient, and the assumption of multicollinearity were met.

CHAPTER FIVE: DISCUSSION

The assessment of teachers' data literacy is an essential contribution to national efforts to strengthen schools' accountability and enhance teachers' use of student data to monitor learning progress. Though various efforts are underway that promotes teachers' use of student data, there is no technically sound measure of teachers' knowledge of student data (Mandinach & Gummer, 2012). This dissertation's main goal is to examine the psychometric properties of that the NU Data Knowledge Scale as a measure of teachers' data literacy.

Research goals

Research Goal 1. To assess the overall consistency of teachers' scores on the NU Data Knowledge Scale instrument.

Research Objective 1a: *To describe the descriptive statistics of the instrument for the sample of teachers****.*** For this research objective, the descriptive statistics of the NU Data Knowledge Scale were examined using traditional methods of scale interpretation (Thorndike & Thorndike-Christ, 2010). A strong measure's descriptive statistics would have an adequate mean and standard deviation in which there was room for two standard deviations of growth above the mean. The NU Data Knowledge Scale had a mean score of 20.77 with a standard deviation of 3.62, suggesting that teachers' knowledge of the databasics varied between participants. The observed mean and standard deviation for this study is important in that it shows that the NU Data Knowledge Scale has an appropriate ceiling for assessing teachers' data literacy over multiple administrations (Ware & Ganderick, 1998). The NU Data Knowledge Scale has a maximum score of 30, so the mean of 20.77 and standard deviation of 3.62 allows for 2.45 standard deviations

of growth. This result confirmed the expectation that the NU Data Knowledge Scale provides an adequate ceiling for scores to grow over time.

Research Objective 1b: To assess the overall consistency of teachers' scores on the instrument. Since the NU Data Knowledge Scale was a single factor, unidimensional scale, the Kuder-Richardson 20 was calculated for the entire instrument. The KR-20 and internal consistency of the teachers' scores on the NU Data Knowledge Scale were assessed using the ltm Package in R. The NU Data Scale had a KR-20 alpha of .6135. A KR-20 value of .70 was chosen as a goal for this measure because it limits the chances of score reversal of high and low achieving participants on retests to 1 in 1000 cases in groups of 25 participants (Thorndike & Thorndike-Christ, 2010). The observed coefficient of .6135, though below the standard level of .70, is still within the acceptable range of reliability for a measure in a new area of research area. Since there is no other published measure of data literacy the observed KR-20 reliability estimate in this study .6135 may provide a preliminary level of internal consistency for future studies of data literacy. This reliability coefficient estimate, though too low for high stakes decisions, is high enough for providing a base-level understanding of teachers' data literacy (Thorndike &Thorndike-Christ, 2010).

Research Objective 1c: To assess the capacity of items to discriminate teachers' knowledge of the databasics. This research objective focused on the discriminative properties of the NU Data Knowledge Scale's items for assessing teachers' data literacy. Traditionally, discriminative items have corrected item-total correlations above .2 with the overall test score. Only 16 items achieved this standard and had a corrected item-total correlation above .2. The average item-total correlation was .19 with a standard deviation

of .11 indicates that there was substantial variance in what the scale items were measuring. Item 10 had a corrected item-total correlation of -.1437 which is indicative of the item fitting the single factor solution and should be revised or deleted from future iterations of measure. Items with low item-total correlations did not contribute to the overall KR-20 and may account for some of the variance in the lower than expected alpha value.

Item-level analyses of test item distractors were expected to reveal negative correlations between each of the items' distractors and the overall score (Frey et al., 2005). Sixteen items had corrected point-biserial correlations above the expected value of .2 for the correct answers. Of these 16 items, only 10 had the correct answer being the only option positively correlated with the overall score. Item 4's distractor B had zero variance due to none of the 215 teachers selecting it and should be revised for future iterations of the measure. The low corrected point-biserial correlations across the scale's items is indicative of having distractors that were too similar to the correct answer that confused the participants. However, an examination of item difficulty revealed that, on average, 69.22 percent of teachers got each item correct.

The apparent disconnect between item difficulty (*p*-value) and the corrected item total correlations for items may be explained by the nature of the NU Data Knowledge Scale. Variations in KR-20 scores, item-total correlations, and item difficulty may be affected by the purpose of the measure and variations in the sample (Thorndike & Thorndike-Christ, 2010). The purpose of the NU Data Knowledge Scale is to assess changes in teachers' knowledge as they progressed through the NU Data intervention. The teachers in this sample were from schools that did not participate in the NU Data

Intervention and have not been exposed to the curriculum. It would be expected that there would be variation in their scores and the measure may have been more internally consistent with teachers who have gone through the NU Data Intervention.

Research Goal 2: To assess the dimensionality and factor structure of the NU Data Knowledge Scale.

Research Objective 2a: To examine the alignment of the subject matter experts' ratings with the test blueprint of the knowledge scale. Two types of analyses were used to determine the alignment of ratings between the three subject matter experts, and the alignment between the aggregate subject matter expert rating and the original NU Data Scale test blueprint. The standard level for adequate subject matter expert absolute agreement would be above 70 percent with unweighted kappa scores above .41 (Vierra & Garret, 2005). The unweighted kappa scores and percentage of absolute agreement for the three subject matter experts in this study achieved these requirements. The percentage of absolute agreement for raters ranged from 79.4 to 81.7, and the unweighted kappa scores ranged from .421 to .455. This is indicative of subject matter experts having an acceptable level of agreement for categorizing items on the NU Data Knowledge Scale by databasics. The level of absolute agreement between the aggregated subject matter expert rating and the original test blueprint was acceptable at 70.6 percent with an unweighted kappa score of .439.

The resulting test blueprint created from the subject matter experts (see Appendix G) was slightly problematic because it reclassified item databasics prior to disbursing the instrument to teachers. As a result, DataBasics 1 (Knows a variety of ways to collect data) and 6 (Selecting evidence-based protocols) were underrepresented in the knowledge

scale. There should be modifications to test items to increase the representation of these databasics in future iterations of the measure. Ideally, an instrument should have equal distributions of test constructs within the exam (Haladyna & Rodriguez, 2013), and each databasic should be represented in at least 20 percent of test items.

Alternatively, databasics could be collapsed into composite databasics in future measures. Means and colleagues (2011) and Mandinach and Gummer (2012) both postulated that it may not be possible to disentangle components of data literacy from each other due to the perceived correlations between their operative elements. Since the NU Data Knowledge Scale is a single-factor construct, the equal representation each of its databasics may not be as critical as once believed, and the underrepresentation of DataBasics 1 and 6 may not be significant.

Research Objective 2b: To determine the dimensionality of the scale through weighted least mean squares (WLMSV) analyses. This research objective examines the dimensionality of the NU Data Knowledge Scale through limited-information weighted least squares mean and variance (WLSMV). The number of factors to retain was determined through the examination of the model fit indices and results from a parallel analysis with 5000 replications. The parallel analysis indicated that a single factor solution should be retained. The resulting fit indices for the single factor solution were mixed. The non-significant Chi-Square and the RMSEA (.019, $p > .999$) supported a single factor solution with this sample of teachers.

Though the single factor solution does not meet all fit indices, it was stronger than the two to six factor solutions and supported by the results of the parallel analyses. The CFI (.870) and TLI (.879), however, were below the recommended level of .9 to indicate

good model fit. Closer examination of the null model's RMSEA's value of .054 indicates that it is below the assumed value of .185 if the null model's CFI and TLI were above .9. This means that the CFI and TLI fit indices are not meaningful fit indices in this particular model, and that the RMSEA and Chi-Square tests are better determinants of model fit (Kenny & McCoach, 2003). Hence, it is believed that data literacy is accurately measured as a single factor construct on the NU Data Knowledge Scale.

Twenty-eight of the 30 items loaded significantly ($p < .05$) on the single factor. The single factor solution accounted for roughly 20 percent of the total variance of item scores overall for the measure. Individual item variances ranged from .071 to .821 and averaged .1890, indicating that the single factor solution accounted for roughly 20 percent of the variance in individual scores. Though the components of data literacy are multifaceted, it has been speculated that a measure of data literacy could be a single factor due to high levels of correlation between the critical components (Means et al., 2011).

Research Goal 3: To describe the relation between scores on the instrument and demographic characteristics of the sample of teachers.

Research Objective 3a: To describe the degree to which years of experience, gender, classroom size, or teaching specialization predict scores on the knowledge scale. This objective was examined by conducting all possible regressions and using Mallow's Cp and adjusted R^2 to determine the best model for predicting teacher scores on the NU Data Knowledge Scale. In total, 1024 different models were analyzed to determine the best combination of factors using the Mallow's Cp (p+1) criterion (Mallows, 1973) and adjusted R^2 scores. Despite the inclusion of 14 possible predictors,

the best fitting models only accounted for roughly 17 percent of the variability in the teachers' overall scores.

This was a surprising finding considering suggestions in previous research. Mandinach and Gummer (2013) suggested that data literacy may be related to years of experience and highest degree received by teachers. Additionally, Dunn and colleagues (2013) postulated that data literacy skills may be directly related to teacher specialization and years of teacher experience. They postulated teachers with more years of education and who specialized in math, special education, or science would be more data literate than generalist teachers. However, the results of this study revealed that years of experience and highest degree received, both accounted for less than one percent of the variance in teachers' overall scores. Means and colleagues (2011) suggested that teachers' confidence in data use may be attributed to their data literacy, but teachers' scores on the Comfort with Data Use survey accounted for commutatively less than 9 percent of the variance in scores on the NU Data Knowledge Scale.

Limitations

There were several limitations in this study. The first limitation is the sample of teachers used in this study. All of the 215 teachers who participated in this study were from rural Nebraska communities outside of the Lincoln/Omaha metroplex. This is problematic in that it limits the representativeness of the sample to the population. On average, the sample in this study had more years of experience, were more highly educated, and were disproportionately more female, and less ethnically diverse than the national population of teachers (U.S. Department of Education, 2013). The differences in the distribution of years of experience, gender, and classroom size between the sample of

teachers in this study and the Nebraska teacher population may be accounted for in the recruitment procedures used in this study. All teachers in this study were recruited from rural districts outside of Omaha and Lincoln. Rural schools may have larger classroom sizes due to budgetary constraints than urban environments, and the sample of teachers may have more years of experience due to difficulty recruiting in these communities. Additionally, the sample of teachers did not participate in the NU Data intervention and, as a result, may not have been exposed to components on the NU Data Knowledge Scale which could have adversely affected the consistency of the overall scale. Indeed, naïve or and teachers with more experience with data-based problem solving, might have demonstrated different response patterns on the measure.

Another weakness of this study was the limitations in the overall consistency of the scale. The KR-20, item-total correlations, and R^2 scores of items within the factor solution were not as strong as expected. The internal consistency, while low, does provide a preliminary internal consistency measure for future research. The single factor solution for this study, though significant did point to some underlying issues. The null data set used for comparison had a RMSEA value below .154 with TLI and CFI below .9. This indicates that the single factor solution may not have been substantially different than the null model.

Another limitation for this study is that the subject matter experts in this study were all academics in educational psychology. Though well versed in educational policy, their experiences within education are different than those of a seasoned teacher working in a classroom. This may have impacted how test items and distractors were refined and

the wording may reflect academic perspectives of teacher knowledge instead of applied teacher knowledge.

Directions for Future Research

Although the NU Data Knowledge Scale needs further refinements, it demonstrates promising utility in measuring teachers' knowledge of data-based decision making, data use, and data literacy. With states and school districts accountable for student achievement and results of high stakes testing, it is important to know the overall data literacy of teachers.

This study provides a preliminary single factor solution measure to be used in future research studies to examine teachers' data literacy. Means and colleagues (2011) outlined the need for two independent measures of data literacy to be developed and compared through convergent reliability. The NU Data Knowledge Scale is now available to be used in future studies and may serve as a tool in exploring the validity evidence of other measures of data literacy. Future studies should examine the correlation between the NU Data Knowledge Scale with other data literacy measures completed by the same sample of teachers. The NU Data Knowledge Scale shows promise of being a measure that blends the skills, data interpretation, and data-based problem solving skills of educators.

One surprising finding of this study is that data literacy appears to be independent of teachers' degree, years of experience, teaching assignment, and comfort with data use. Future studies should focus on examining the relation between teacher in-service trainings and their data literacy. This would provide a more localized focus of teacher data use. The measure could be useful in conducting needs assessments of school

districts, and could possibly be used as a pre/post measure of teacher training and preparation programs.

This study also provided a new model for measure development that incorporates item banks, subject matter experts, and rigorous analysis of the psychometric properties of a measure. Future studies could examine teachers who completed the NU Data intervention to see if their scores are substantially different than teachers who did not participate in the NU Data intervention. This study could highlight differences in how teacher training programs affect data literacy and if that affects the internal reliability and item-total correlations of the measure.

Follow-up studies may want to reexamine the test items and distractors with a subject matter expert panel of seasoned teachers. This would facilitate discussion between academics and teachers about the types of items and the content of items included in the measure. The resulting revisions to the measure and effect on the psychometric properties of the instrument could be meaningful to expanding the utility of the measure.

Implications for Practice

There are three implications for future practice from this study. First, other researchers may be interested in using the methods of measure development employed in this study in future projects. This would enable other reserachers to develop measures with strict psychometric properties. Other researchers examining data literacy could use the NU Data Knowledge Scale as a referece criterion, or as a convergent measure of data literacy. This would provide the field with a launching point for future data literacy research.

This study explored the relations between years of expereince, size of the classroom, teacher speicialization and assignment, and highest degree received and found that they were not practically significant in predicting teacher data literacy. As a result, it would be beneficial for future resereachers to examine non-traditional sources of teacher data literacy. Future data literacy research should examine the relation of environmental influences of teacher data literacy such as in-service training programs, administrative attitude towards data use, and the use of data coaches within schools. This would allow reseachers to expand upon the results of this study and strengthen data literacy research.

Additionally, the results of this study highlight discpreancies in teacher knowledge of data use. In order to assess teachers' data literacy, there needs to be stronger foresight and training standards at the fedral level that mandate teacher training in effective data use. Without federal mandates in teacher training, these discrepancies in teachers' data literacy will persist as teacher training programs can pick and choose which components of data use they want to teach. One possible solution to addressing these discrepancies in data literacy is to provide inservice opportunites for teachers to address weaknesses in their knowledge of data use.

Finally, the results of this study warrant several revisions to the measure in response to analyses. Items on the measure should be adjusted to reflect DataBasic 1 and DataBasic 6 so that each databaic is reflected in at least 20 percent of test items. Additionally, item 10 should be revised since it detracts from the KR-20 and did not fit the single factor solution. Items 18 and 26 should also be revised as they did not fit the sinlg factor solution either. Revising items 10, 18, and 26 could help balance the distribution of databasics in the measure. These changes to the NU Data Knowledge

Scale may strengthen the psychometric properties and dimensionality of the measure in future iterations.

References

Abbott, D. V. (2008). A functionality framework for educational organizations: Achieving accountability at scale. In E. B. Mandinach & M. Honey (Eds.), *Data driven school improvement: Linking data and learning* (pp. 257-276). New York, NY: Teachers College Press.

Anderson, S., Leithwood, K., & Strauss, T. (2010). Leading data use in schools: Organizational conditions and practices at the school and district levels. *Leadership and Policy in Schools, 9,* 292-327. doi: 1.1080/15700761003731492

Bakx, A., Baartman, L., & van Schilt-Mol, T. (2014). Development and evaluation of a summative assessment program for senior teacher competence. *Studies in Educational Evaluation, 40,* 50-62. doi: http://dx.doi.org/1.1016/j.stueduc.2013.11.004

Bautista, A., Cañadas, M. C., Brizuela, B. M., & Schliemann, A. D. (2015). Examining how teachers use graphs to teach mathematics during a professional development program. *Journal of Education and Training Studies*, 3, 91-106.

Bernhardt, V. (2013). *Data, data, everywhere: Bringing all the data together for continuous school improvement*: Routledge.

Booher-Jennings, J. (2005). Below the bubble:"Educational triage" and the texas accountability system. *American Educational Research Journal, 42,* 231-268.

Boudett, K. P., City, E., & Murnane, R. (2005). *Data wise: A step-by-step guide to using assessment results to improve teaching and learning*. Cambridge, MA: Harvard Education Press.

Brown, T. (2006). Confirmatory factor analysis for applied research. New York: Guildford.

Brunner, C., Fasca, C., Heinze, J., Honey, M., Light, D., Mardinach, E., & Wexler, D. (2005). Linking data and learning: The grow network study. *Journal of Education for Students Placed at Risk, 10*, 241-267.

Choppin, J. (2002). *Data use in practice: Examples from the school level.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P. M. (2014). Searching for stages of teacher's skills in assessment. *Studies in Educational Evaluation, 40*, 1-11. doi: http://dx.doi.org/1.1016/j.stueduc.2013.11.006

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis* (2nd ed.). Hillside, NJ: Lawrence Erlbaum Associates, Inc. .

Costello, A. B., & Osborne, J. W. (2005). Getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*, 1-9.

Council for the Accredidation of Educator Preparation. (2013). Caep accredidation standards. Retrieved from http://caepnet.org/ website:

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302. doi: 1.1037/h0040957

Doll, B., Franta, E., Thomas, A., Chapla, B., & Sikorski, J. (2014, April). *Nu data: Preparing educators to use data well*. Paper presented at the American Educational Reserach Association, Philadelphia, PA.

Doll, B., Haack, K., Kosse, S., Osterloh, M., Siemers, E., & Pray, B. (2005). The dilemma of pragmatics: Why schools don't use quality team consultation practices. *Journal of Educational and Psychological Consultation, 16*, 127-155. doi: 1.1207/s1532768xjepc1603_1

Doll, B., Horn, C., & Shope, R. (2010). Nu data: Using data to foster the school success of students with disabilities (pp. 1-25): Institute of Educational Sciences.

Dunn, K. E., Airola, D. T., Lo, W.-J., & Garrison, M. (2013). Becoming data driven: The influence of teachers' sense of efficacy on concerns related to data-driven decision making. *Journal of Experimental Education, 81*, 222-241. doi: 1.1080/00220973.2012.699899

Frey, B. B., Petersen, S., Edwards, L. M., Pedrotti, J. T., & Peyton, V. (2005). Item-writing rules: Collective wisdom. *Teaching and Teacher Education, 21*, 357-364. doi: http://dx.doi.org/1.1016/j.tate.2005.01.008

Gettinger, J. B., Mulford, L., & Hoffman, A. (2010). Prevention and early intervention for preschool children at risk for learning and behavior problems. In Doll, W. Pfohl & J. Yoon (Eds.), *Handbook of youth prevention science* (pp. 349-374). New York: Routledge.

Goldring, E., & Berends, M. (2008). *Leading with data: Pathways to improve your school*: SAGE.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of personality assessment, 68*, 532-56.

Gulikers, J. T. M., Biemans, H. J. A., Wesselink, R., & van der Wel, M. (2013). Aligning formative and summative assessments: A collaborative action research

challenging teacher conceptions. *Studies in Educational Evaluation, 39*, 116-124. doi: http://dx.doi.org/1.1016/j.stueduc.2013.03.001

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*, 309-333. doi: 1.1207/S15324818AME1503_5

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York, NY: Routledge.

Halverson, R., Prichett, R. B., & Watson, J. G. (2007). *Formative feedback systems and the new instructional leadership*. Madison, WI: Wisconsin Center for Eduational Research

Hansen, J. D., & Dexter, L. (1997). Quality multiple-choice test questions: Item-writing guidelines and an analysis of auditing testbanks. *Journal of Education for Business, 73,* 94-97. doi: 1.1080/08832329709601623

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods, 7*, 191-205.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*, 1-55

Huguet, A., Marsh, J. A., & Farrell, C. (2014). Building teachers' data use capacity: Insights from strong and developing coaches. *education policy analysis archives, 22*, 52.

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106,* 1258-1287. doi: 1.1111/j.1467-962.2004.00379.x

Jacobs, J., Gregory, A., Hoppey, D., & Yendol-Hoppey, D. (2009). Data literacy: Understanding teachers' data use in a context of accountability and response to intervention. *Action in Teacher Education, 31,* 41-55. doi: 1.1080/0162662.2009.10463527

Jimerson, J. B. (2014). Thinking about data: Exploring the development of mental models for "data use" among teachers and school leaders. *Studies in Educational Evaluation, 42,* 5-14. doi: 1.1016/j.stueduc.2013.1.010

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*: Oxford University Press.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. Structural equation modeling, 10, 333-351.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th Ed.). Sanndle River, NJ: Parson: Prentice-Hall, Inc.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education, 112,* 496-552.

Knapp, T. R., & Mueller, R. O. (2010). Reliability and validity of instruments. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewers guide to quantitative methods in the social sciences* (pp. 337-341). New York, NY: Routledge.

Kosse, S., & Doll, B. (2006, March). A comparison of traditional student assistance teams and response to intervention. A paper presented at the Annual Convention of the National Association of School Psychologists, Anaheim, CA.

Lai, M. K., & Hsiao, S. (2014). Developing data collection and management systems for decision-making: What professional development is required? *Studies in Educational Evaluation*, *42*, 63-7. doi: http://dx.doi.org/1.1016/j.stueduc.2013.12.006

Lawshe, C. H. (1975). A quantitative approach to content validity1. *Personnel psychology*, *28*, 563-575.

Lieberman, G. M., & Morris, J. D. (2008). Mallow's cp for selecting best performing logistic regression subtests. *Mulitple Linear Regression Viewpoints*, *34*, 8-12.

Long, L., Rivas, L., Light, D., & Mandinach, E. (2008). The evolution of a homegrown data warehouse: Tusdstats. In E. B. Mandinach & M. Honey (Eds.), *Data-driven school improvement: Linking data and learning* (pp. 209-232). New York, NY: Teachers College Press.

Love, N. (2004). Taking data to new depths. *Journal of Staff Development*, *25*, 22-26.

Love, N. (2011). Data literacy for teachers: Dude Publishing.

Love, N., Stiles, K. E., Mundry, S., & DiRanna, K. (2008). *The data coach's guide to improving learning for all students: Unleashing the power of collaborative inquiry*. Thousand Oaks, CA: Corwin Press.

Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, *23*, 26-32. doi: 1.1111/j.1745-3992.2004.tb00156.x

Mallows, C. L. (1973). Some comments on c p. *Technometrics, 15,* 661-675.

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist, 47,* 71-85. doi: 1.1080/0046152.2012.667064

Mandinach, E. B., & Gummer, E. (2012). Navigating the landscape of data literacy: It is complex. *Washington, DC and Portland, OR: WestEd and Education Northwest*.

Mandinach, E. B., & Gummer, E. S. (2013a). Building educators' data literacy: Differing perspectives. *The Journal of Educational Research & Policy Studies, 13,* 4-8.

Mandinach, E. B., & Gummer, E. S. (2013b). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42,* 30-37. doi: 1.3102/0013189x12459803

Mandinach, E. B., Honey, M., & Light, D. (2006). *A theoretical framework for data-driven decision making.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391-410

Maslow, A. (1962). Toward a psychology of being. Princeton, NJ: D. Van Nostrana CoInc Inc.

McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., . . . Gray, A. L. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities, 35*, 69-86. doi: 1.1177/002221940203500106

Means, B., Chen, E., DeBarger, A., & Padilla, C. (2011). Teachers' ability to use data to inform instruction: Challenges and supports *Office of Planning, Evaluation and Policy Development, US Department of Education* (pp. 1-122). Washington, DC: US Department of Education.

Mertler, C. A., & Campbell, C. (2003). *Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference.* Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH.

Menard, S. (1995). Applied logistic regression analysis. In *Sage University Paper Series on Quantitative Applications in the Social Sciences No. 106*. Thousand Oaks, CA: Sage.

Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge & application of classroom assessment concepts: Development of the assessment literacy inventory*. Paper presented at the American Educational Research Association, Montreal, Quebec, Canada.

Myers, R. H. (1990). *Classical and Modern Regression Application (2nd Edition)*. Pacific Grove, CA.

Newton, J. S., Algozzine, B., Algozzine, K., Horner, R. H., & Todd, A. W. (2011). Building local capacity for training and coaching data-based problem solving with positive behavior intervention and support teams. *Journal of Applied School Psychology, 27*, 228-245. doi: 1.1080/15377903.2011.590104

Pedhauzer, E. J. (1997). *Multiple-regression in behavioral research: Explanation and prediction* (3rd ed). New York, NY: Harcourt Brace. DOI: 10.2307/2285468

Penuel, W. R., Roschelle, J., & Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. *Research and Practice in Technology Enhanced Learning, 2*1, 51-74.

Preacher, K. J., & Coffman, D. L. (2006, May). Computing power and minimum sample size for RMSEA [Computer software]. Available from http://quantpsy.org/.

Qin, J., & D'Ignazio, J. (2010). *Lessons learned from a two-year experience in science data literacy education*. Paper presented at the International Association of Scientific and Technological University Libraries, 31st Annual Conference, Purdue University.

Quilter, S. M., & Gallini, J. K. (2000). Teachers’ assessment literacy and attitudes. *The Teacher Educator, 36*, 115-131.

Reschly, A. L., & Christenson, S. L. (2006). Prediction of dropout among students with mild disabilities a case for the inclusion of student engagement variables. *Remedial and Special Education, 27*, 276-292.

Roehrig, A. D., Duggar, S. W., Moats, L., Glover, M., & Mincey, B. (2008). When teachers work to use progress monitoring: Data to inform literacy instruction identifying potential supports and challenges. *Remedial & Special Education, 29*, 364-382.

Schaughency, E., & Ervin, R. (2006). Building capacity to implement and sustain effective practices to better serve children. *School Psychology Review, 35*, 155-166.

Schield, M. (2004). Information literacy, statistical literacy and data literacy. *IASSIST Quarterly, 28*, 6-11.

Schoenfeld, A. H. (2007). The complexities of assessing teacher knowledge. *Measurement: Interdisciplinary Research and Perspectives, 5*, 198-204. doi: 1.1080/15366360701492880

Sikorski, J., Doll, B., Thomas, A., Franta, E., & Kenney, C. (2013). Nu data: Building educators' data use in schools. *Educational Researcher*, 23-27.

Sikorski, J., Franta, E., & Doll, B. (2014). *Measuring educators' data literacy: The nu data knowledge scale*. Paper presented at the National Association of School Psychologist Annual Convention, Washington, DC.

Sireci, S. (1998). The construct of content validity. *Social Indicators Research, 45*, 83-117. doi: 1.1023/A:1006985528729

Smith, K. (2013). Formative assessment of teacher learning: Issues about quality, design characteristics and impact on teacher learning. *Teachers & Teaching, 19*, 172-178. doi: 1.1080/13540602.2013.741835

Staman, L., Visscher, A. J., & Luyten, H. (2014). The effects of professional development on the attitudes, knowledge and skills for data-driven decision making. *Studies in Educational Evaluation, 42*, 79-9. doi: http://dx.doi.org/1.1016/j.stueduc.2013.11.002

Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Stephens, D., Pearson, P. D., Gilrane, C., Roe, M., Stallman, A. C., Shelton, J., . . . Commeyras, M. (1995). Assessment and decision making in schools: A cross-site analysis. *Reading Research Quarterly, 30*, 478-499.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*, 238-238.

Suen, H. K., Lei, P.-W., & Li, H. (2011). Data analysis for effective decision making. In M. A. Bray & T. J. Kehle (Eds.), *The oxford handbook of school psychology* (pp. 140-168). Oxford, NY: Oxford University Press.

Sun, W., Chou, C.-P., Stacy, A. W., Ma, H., Unger, J., & Gallaher, P. (2007). Sas and spss macros to calculate standardized cronbach's alpha using the upper bound of the phi coefficient for dichotomous items. *Behavior research methods, 39*, 71-81.

Tabachnick, B. G., and Fidell, L. S. (2013). *Using multivariate statistics, 6th Ed.* Boston, MA: Pearson.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*: American Psychological Association.

Thorndike, R. L., & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education* (8th ed.). Boston, MA: Pearson.

U.S. Department of Education. (2013). Digest for education statistics. Table 209.3. from https://nces.ed.gov/programs/digest/d13/tables/dt13_209.3.asp

U.S. Department of Education. (2008). Teachers' use of student data systems to improve instruction: 2005 to 2007.

Vahey, P., Rafanan, K., Patton, C., Swan, K., Hooft, M., Kratcoski, A., & Stanford, T. (2012). A cross-disciplinary approach to teaching data literacy and

proportionality. *Educational Studies in Mathematics, 81*, 179-205. doi: 1.1007/s10649-012-9392-z

Vahey, P., Rafanan, K., Patton, C., Swan, K., van't Hooft, M., Kratcoski, A., & Stanford, T. (2012). A cross-disciplinary approach to teaching data literacy and proportionality. *Educational Studies in Mathematics, 81*, 179-205.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Fam Med*, *37*, 360-363.

Ware Jr, J. E., & Gandek, B. (1998). Methods for testing data quality, scaling assumptions, and reliability: The iqola project approach. *Journal of Clinical Epidemiology, 51*, 945-952. doi: http://dx.doi.org/1.1016/S0895-4356(98)00085-7

Wayman, J. C., & Jimerson, J. B. (2014). Teacher needs for data-related professional learning. *Studies in Educational Evaluation*, *42*, 25-34. doi: http://dx.doi.org/1.1016/j.stueduc.2013.11.001

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*, 806-838. doi: 1.1177/0011000006288127

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research, 25*, 508-518. doi: 1.1177/0193945903252998

Ysseldyke, J., Burns, M., Dawson, P., Kelley, B., Morrison, D., Ortiz, S., . . . Telzrow, C. (2009). The blueprint for training and practices as the basis for best practices. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5 ed., Vol. 1, pp. 37-70). Bethesda, MD: NASP.

APPENDIX A

Teacher Consent Form
Project: Knowledge Scale Item Development
Purpose of the research: The purpose of the Knowledge Scale Item Development project is to develop a measure that assesses teachers' knowledge of data collection and measurement. You are invited to participate because you are a teacher working within the Midwest. You will spend between twenty to thirty minutes working with the project.
Procedures: If you agree to participate in this study, you will be asked to complete the following measures. The knowledge scale will ask you to indicate some descriptive information such as: gender; years of teaching experience; and specialization. An addition attitudes survey will ask you to rate you beliefs on data use. The measures will require twenty to thirty minutes complete.
Risks and/or discomforts: There are no known risks associated with participation in this study.
Benefits: You will benefit from taking this measure by reviewing your knowledge of data collection and measurement. The university will benefit from the development of a new knowledge scale measure for strengthening educator-training programs.
Confidentiality: No identifying information will be collected. Records from this study will be stored for 5 years in a locked cabinet in a Center for Instructional Innovation office and will only be seen by the researchers or research assistants. The information obtained in this research may be published in scientific journals or presented at professional meetings, but the data will be reported as group data and will not identify any participant. An email address will be collected at the discretion of the teacher in order to enter the drawing for the five 50 dollar gift cards. These email addresses will not be linked to any participant responses.
Compensation: Teachers are eligible for a drawing of five 50 dollar Amazon gift cards at the end of the data collection. The overall odds of receiving the gift card depend on the number of participants, but you have at least a 1 in 60 chance of receiving a gift card.
Opportunity to ask questions: You can ask questions about this study and get answers to your questions now or anytime during the research. Or, you may call Doll 402.472.2238 or email bdoll2@unl.edu, or call Jonathon Sikorski 402.45.1549 or email jonathon.sikorski@huskers.unl.edu to discuss the research. If you have concerns about the study or if Dr. Doll and Mr. Sikorski cannot answer questions about your rights as a research teacher, you can contact the University of Nebraska Lincoln Institutional Review Board at 402.472.6965.
Freedom to withdraw: Your participation in this study is voluntary. You are free to decide not to participate or to withdraw at any time. If you withdraw from the study, it will not harm your relationship to Dr. Doll, Jonathon Sikorski, or the University of Nebraska. If you withdraw, you will not lose any benefits that you are otherwise entitled to receive.

Consent: By continuing this survey, you are saying that you agree: (1) to participate in the NU Data Item Development Study, (2) that you currently an educator working within a school, and (3) that you have read and understood the information above. Please print or save a copy of this consent form for your own records.
Jonathon Sikorski MS. E. Principal Investigator, Office: 402 450 1549
Q39 I have read the consent form and agree to participate in the research study.
□□Agree (1)
□□Disagree (2)

If Disagree Is Selected, Then Skip To End of Survey

APPENDIX B
Demographics

Q39 I have read the consent form and agree to participate in the research study.
- ❍ Agree (1)
- ❍ Disagree (2)

If Disagree Is Selected, Then Skip To End of Survey

Q31 Gender
- ❍ Male (1)
- ❍ Female (2)
- ❍ Other (3)

Q33 Highest Degree Received
- ❍ High school/GED (1)
- ❍ Some secondary education (2)
- ❍ Associates Degree (3)
- ❍ Bachelors Degree (4)
- ❍ Masters Degree (5)
- ❍ Education Specialist Degree (or Masters +30) (6)
- ❍ Doctoral Degree (PhD or EdD) (7)

Q38 Number of years of working in educational settings:
- ❍ 1 to 5 years (1)
- ❍ 6 to 10 years (2)
- ❍ 11 to 15 years (3)
- ❍ 16+ years (4)

Q33 What is your current specialization in schools?
- ❍ General education
- ❍ Special education
- ❍ Speech/language
- ❍ Other

Q35 How many students are in your classroom?
- o 0 to 4
- o 5 to10
- o 11 to 15
- o 16 to 20
- o 21 to 25
- o 26 to 30
- o 31+ (or you are specialist/rotate classes)

APPENDIX C

NU Data Knowledge Scale

Q1 When is an intervention evidence based?
- ❍ a. When other people have used it and found that it worked well (1)
- ❍ b. When someone has collected pre and post data that shows meaningful change (3)
- ❍ c. When the author says that the intervention shows meaningful change (2)
- ❍ d. When it is published by a respected professional organization (4)

Q2 What is baseline data?
- ❍ a. Data that is collected over a short period of time (1)
- ❍ b. Data that is collected after an intervention is implemented (2)
- ❍ c. Data collected before an intervention has been implemented (3)
- ❍ d. The bottom 25 percent of all data points (4)

Q3 Which of the following is an example of a strong goal statement?
- ❍ a. The student's on-task behavior will improve by the end of the semester according to teacher report. (2)
- ❍ b. The student's fluency will improve significantly over the semester as measured by a curriculum based assessment (3)
- ❍ c. The student's grades will be a "B" or better by the end of the semester according to report cards (4)
- ❍ d. The student will increase math computation scores by 15 points by the end of the semester as measured by curriculum based assessments (1)

Q4 Which of the following is an observable behavior?
- ❍ a. The number of times a student blurts out in class (2)
- ❍ b. A student's attitude towards math or science class (1)
- ❍ c. The intensity of a student's feelings about a poor grade (3)
- ❍ d. The number of times a student becomes frustrated (4)

Q5 How would you translate a student's behavior ratings of Rarely, Sometimes, Often, or Almost Always into data that could be used in a graph?

- ❍ a. By counting the number of times each rating is used to describe the student's behavior (1)
- ❍ b. By comparing the first day's rating to the most recent day's rating (2)
- ❍ c. By assigning a numerical value to each behavior rating (3)
- ❍ d. By grouping together the positive ratings in one graph and the negative ratings into a second graph (4)

Q6 Which of the following best describes an effective progress monitoring strategy?

- ❍ a. Weekly average scores for 20-word quizzes in English class (5)
- ❍ b. The daily number of words read correctly on selected reading passages (7)
- ❍ c. Total scores on a student’s mid-term and final in Algebra (6)
- ❍ d. Standardized test scores of a student’s cognitive abilities (8)

Q7 Your kindergarten team is working with a student who is struggling with pre-literacy skills. Which of the following is an acceptable strategy for strengthening the student's pre-literacy skills?

- a. Playing a pre-literacy game that a colleague developed
- b. Retaining the student for an additional year to help the student better understand the curriculum
- c. Changing the student's seat assignment to put the student closer to the teacher
- d. Implementing an early literacy intervention that you found in a peer reviewed manual

Q8 In addition to class grades, how could you reliably measure the academic performance of students in reading?

- a. By asking the students how they are doing (2)
- b. By reviewing curriculum based measures benchmark scores (1)
- c. By reviewing their academic grades from the last two year (3)
- d. By giving the students a pop-quiz in one of their classes (4)

Q9 What should you do before collecting information on a student?
- a. Collect baseline data on the student in his/her core classes (1)
- b. Decide how you are going to graph the data that you collect (2)
- c. Define and describe the identified problem (3)
- d. Meet with the student to talk about the identified problem (4)

Q10 What is an important consideration when choosing an intervention?
- a. How easily is the intervention implemented (1)
- b. How will the data look once it is graphed (2)
- c. How popular is the intervention package in my district (3)
- d. How many students does the intervention affect (4)

Q11 You have been monitoring the number of times a student was out of seat during class. Your data show that the student has good days and bad days but it is hard to tell if the student is improving. How could a graph show the parents whether the student is making progress?

- a. Create a line graph that compares the student's out-of-seat behavior to that of a typical classmate (1)
- b. Graph the data and add a line that separates the before intervention and during intervention data (3)
- c. Collapse the data into weekly averages and graph them to decide whether these decrease over time (4)
- d. Graph the data and add a trend line that represents the trend in the student's out of seat behavior (2)

Q12 Pat is constantly disrupting class by being out of his seat. How would you measure how much Pat was out of his seat during an observation?

- a. Putting a hash mark on a piece of paper every time he left his seat (2)
- b. Starting a timer every time he leaves his seat and stopping it when he returns (1)
- c. Rating the disruptiveness of his out of seat behavior on a scale from 1 to 10 (3)
- d. Counting the number of students who stopped working when he left his seat (4)

Q13 Your team has collected data on a student with behavioral disorders for several weeks and is now ready to implement an intervention. How would you show where an intervention started on a line graph?

- a. Label where the intervention began on the data file worksheet (1)
- b. Create a separate graph for the pre-intervention and post intervention data (2)
- c. Draw a vertical line on the graph that separates the baseline data from the intervention data (3)
- d. Draw a line that shows the trend across pre and post intervention data (4)

Q14 A teaching team is worried about a student who is not passing her English class. The team's data show that the student increased her work completion from 45% of assignments to 60% of assignments. Still, the student was failing because her grades on each assignment were still low. Given what the team knows, what would be their next step?

- a. Gather data on work accuracy in addition to work completion (4)
- b. Shorten the length of the student's assignments (1)
- c. Place the student on a behavior plan (2)
- d. Gather data on the student's attention to work (3)

Q15 A third grade teacher surveyed the students to see which subject was their favorite: math, science, reading, or social studies. If the teacher wanted to show the student's a graph describing the percent of students preferring each subject, which graph should they choose?

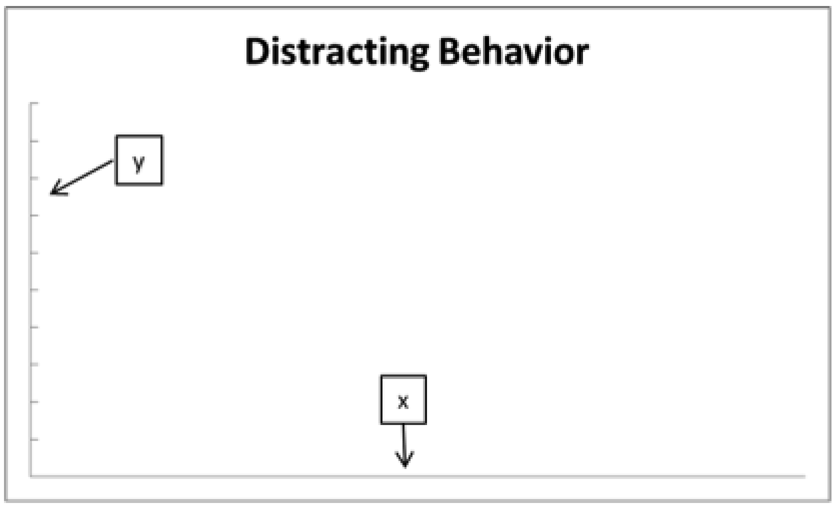
- a. Stacked bar graph (1)
- b. Pie Graph (4)
- c. Line graph (2)
- d. Scatter plot (3)

Q16 A sixth grade teacher notices that her students appear to be arguing and complaining more than usual. She would like to collect data about what is taking place in her

classroom. What kind of data collection would be useful in collecting the information she wants to measure?

- a. Tally the number of times students argued or complained in her classroom (2)
- b. Give the students a survey asking which subjects are their favorite at school (3)
- c. Consult fellow teachers for their thoughts and ideas on ways to improve the classroom climate (4)
- d. Collect anonymous survey data on the students' perceptions of the classroom environment (1)

Q17 After collecting data on a student's distracting behavior in class, a teacher wants to graph the number of times the student engaged in distracting behavior over a two week period on the line graph below. What would you label the x and y axes in the below graph?



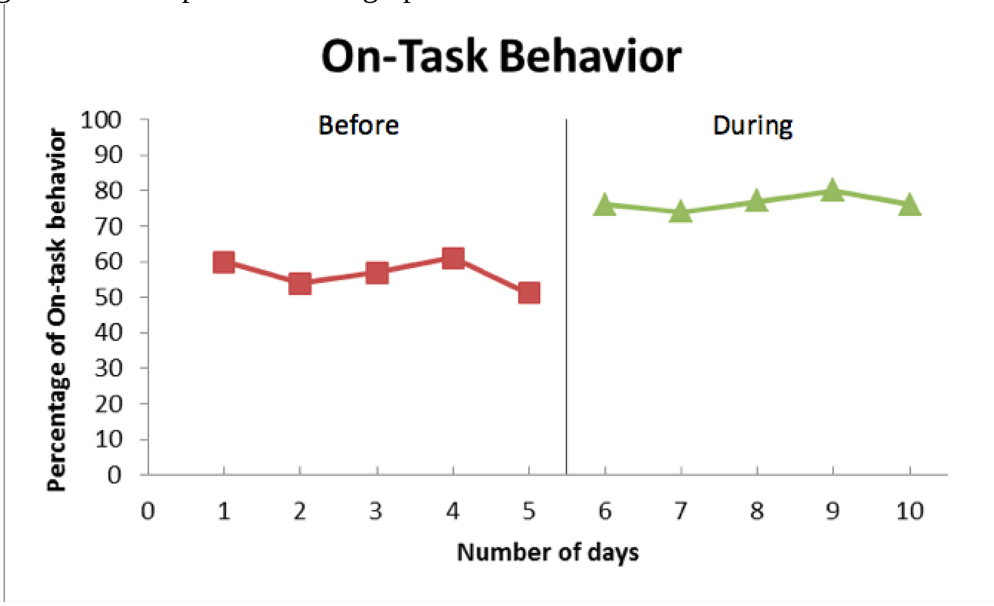
- a. The number of behaviors on the x-axis and the class periods on the y-axis (1)
- b. The date on the x-axis and the number of distracting behaviors on the y-axis (2)
- c. The weekly averages on the x-axis and the number of distracting behaviors on the y-axis (3)
- d. The date on the x-axis and the number of times the student left his seat on the y-axis (4)

Q18 A special education team met to make a plan for a student with a behavior problem. They defined the target behavior and created a rating scale from 0 to 4; with 0 representing a bad day and 4 representing a good day. They set a goal of the student earning a 3 or better each day. They collected data, but when they graphed it they were

disappointed by how far the student was from meeting the goal they set. What step did the team forget in creating their plan?

- a. Collect data describing the student's initial behavior (1)
- b. Rate the student's behavior in multiple settings (3)
- c. Define a goal related to the student's behavior (2)
- d. Record the student's behavior in a continuous manner (4)

Q19 How could you describe the difference in data between the before intervention and during intervention phases on the graph below?



- a. It is impossible to tell because there was too much variability in each phase (1)
- b. The intervention was not effective because the student became less on-task after the intervention was implemented (2)
- d. The intervention was effective in increasing the student's on-task behavior because the data in the during phase do not overlap with data in the before phase (3)
- c. There are not enough data points to determine if the intervention had an effect (4)

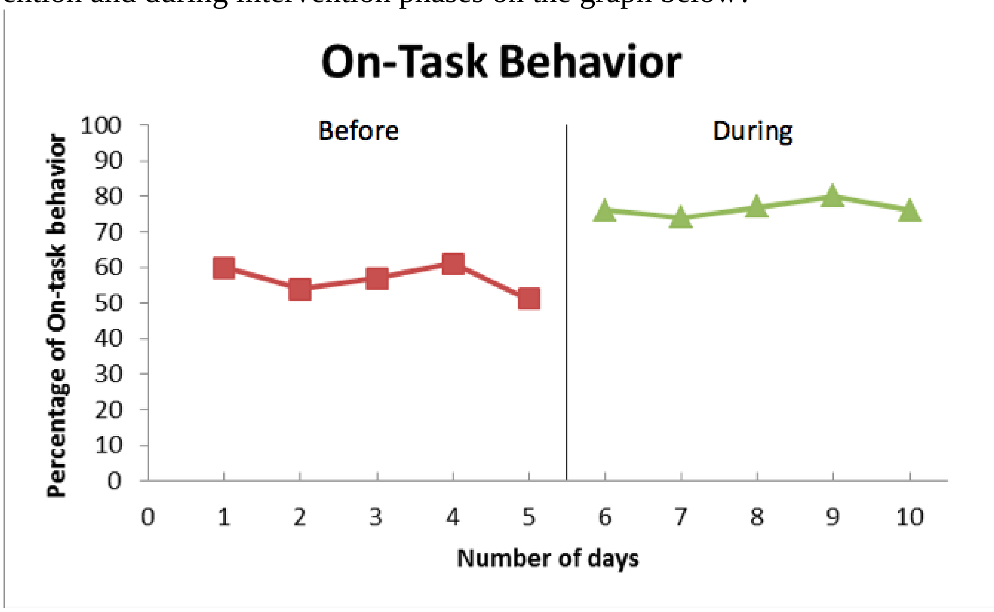
Q20 A student with cognitive disabilities is being taught the steps of washing her hands. She routinely skips one or two steps and becomes frustrated. To figure out which steps of hand washing the students was skipping, what should the teacher do next?

- a. Count every time the student leaves to wash her hands and record it at the end of the day (2)
- b. Put a tally mark on a piece of paper every time the student washes her hands correctly (3)
- c. Rate how well the student washed her hands on a scale from 1 to 10 and graph her daily percentages (4)
- d. Write the hand washing steps in order and circle which ones she completes (1)

Q21 Your team was referred a student who has a history of being extremely difficult and resistant with frequent tantrums. Your team is not sure why the problem behavior occurs. What should they do?

- a. Use a broad screening measure to gather data on behaviors, academic skills, participation, and learning for the student (4)
- b. Implement an evidence based intervention program that reduces the tantruming behavior (1)
- c. Call an IEP meeting to determine whether to qualify the student for special education services (2)
- d. Observe the student at lunch and recess and tally the number of times the student tantrums (3)

Q22 A teacher has been implementing an academic intervention to increase a student's test scores. How should you describe the difference in data between the before intervention and during intervention phases on the graph below?



- a. The intervention was effective in increasing the student's quiz scores because data in the during phase do not overlap with data in the before phase (2)
- b. The intervention was not effective in increasing the student's quiz scores because there was not a significant difference across phases (1)
- c. The intervention was not effective because the student's quiz score decreased after the intervention was implemented (3)
- d. There are not enough data points in before and during phases to determine if the intervention had an effect (4)

Q23 You collected 5 days of baseline data and 5 more days of data after beginning an intervention. You graphed the data on a line graph but are unable to tell whether the

student’s on-task behavior improved. What could you do to clarify changes in the student’s on-task behavior over time?

- a. Draw a horizontal line on the graph to show your goal for the student’s behavior (1)
- b. Draw a trend line on the graph to show the 10 day pattern of the student’s on-task behavior (3)
- c. Separate baseline and intervention data and draw a trend line for each phase (2)
- d. Change the graph type to bar graph (4)

Q24 A team of four teachers met briefly to create a plan for collecting data on a student who was consistently disruptive during their classes. They all decided to record the number of times the student was off-task during their class periods by tallying the number of times the student was disruptive. When they met after school, their tallies varied greatly and they could not agree on what the student’s problem behavior was. What did they forget to do before collecting data on the student?

- a. Define the target behavior in precise terms (1)
- b. Observe the student during recess (2)
- c. Have frequent meetings about the student (3)
- d. Decide how often they would tally the behaviors (4)

Q25 You have been collecting data on a student for several weeks and decided to implement an intervention with the goal that your student would increase work completion from 45 percent to 80 percent of assignments. Your student has not missed completing an assignment for the last three weeks and appears to have reached this goal. What should you do next?

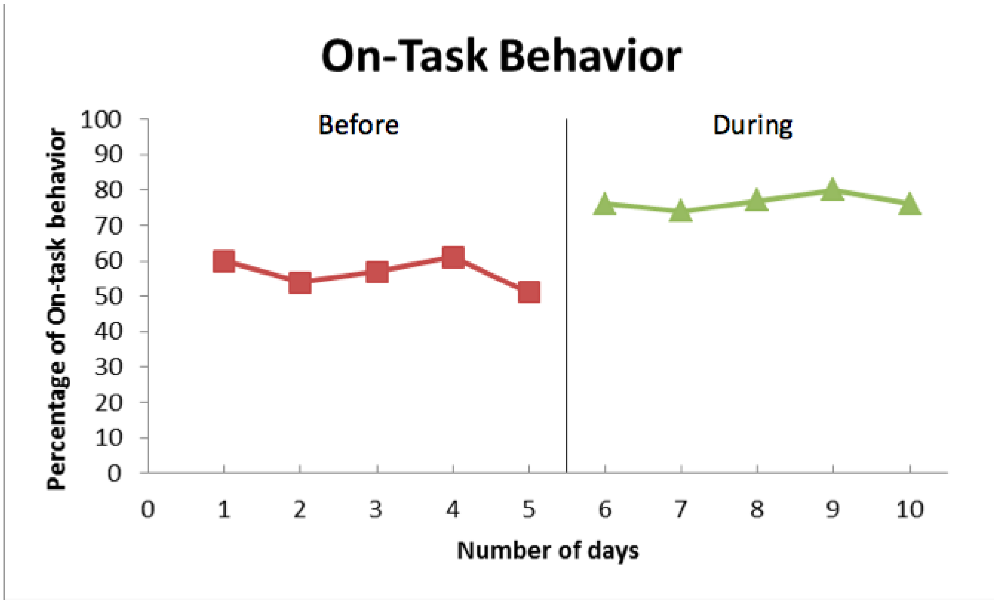
- a. Continue the intervention and data collection (1)
- b. Discontinue collecting and recording the student’s progress (2)
- c. Implement an additional interventions that supports work completion (3)
- d. Decrease the intervention and continue to monitor the student’s work completion for two more weeks (4)

Q26 A middle school team has been implementing an evidence-based intervention for student bullying. They are concerned that the teasing program might not be appropriate because 90% of their school’s enrollment is Latino/a and the program was written for a low-income, predominately white college town in Southern Mississippi. Parents requested the school use an evidence-based program, but a counselor urges the team to

use a peacemaking program that was written by a former counselor in the district because it is a better match to the students' culture. What should the team do?

- ❍ a. They should adapt the evidence-based program to fit the culture of their students (4)
- ❍ b. They should choose the evidenced based intervention because it has been empirically tested (1)
- ❍ c. They should choose the peacemaking program because it was designed for the culture of the district (2)
- ❍ d. They should drop the peacemaking program because it is not evidenced based (3)

Q27 How would you describe the trend test scores in the before and during phases of the graph below?



- ❍ a. The before trend is increasing and the during trend is decreasing (1)
- ❍ b. Both the before and during trend are increasing (2)
- ❍ c. The before trend is decreasing and the during trend is increasing (3)
- ❍ d. The before trend is improving and the during trend is worsening (4)

Q28 An Art teacher wanted to record how much she was praising a First Grade classroom for positive behavior during a lesson. How could the teacher measure how much she praised the students?

- a. Starting a timer every time a student was praised and stopping it when the teacher had to redirect a student (2)
- b. Putting a hash mark on a piece of paper every time the teacher praised students and adding the total praises at the end of the lesson (1)
- c. Rating how well the teacher praised students on a scale from 1 to 10 and graphing the results (3)
- d. Counting the number of students the teacher had a positive interaction with and recording the total (4)

Q29 Midway through the year, a newly-enrolled third grade student is referred to your student assistance team because of an inability read at grade level. How could your team gain the most useful information about this student's current reading abilities and instructional needs?

- a. Conduct observations of the student during reading instruction (1)
- b. Give the student a standardized achievement battery and cognitive abilities test (2)
- c. Collect information about the student's quarterly grades in all academic areas (4)
- d. Use curriculum based measures to track the student's progress in reading (3)

Q30 You are working with a first year teacher to use a behavioral intervention program to reduce classroom interruptions and teach cooperative work behaviors. Your role is to coach the teacher in using the intervention and provide the materials to implement it properly. After two weeks the teacher stopped the intervention because it was not working. What should you do next?

- a. Observe the teacher during a class to see if the intervention was implemented as it was written (1)
- b. Discontinue the intervention and try another evidence-based program (2)
- c. Refer the problem students for special educational services because they are resistant to intervention (3)
- d. Add additional behavioral interventions to the classroom to see if these improve classroom behavior (4)

Q40 Would you like to be entered into a drawing for one of three $150 amazon gift cards?

- Yes (3)
- No (4)

If No Is Selected, Then Skip To End of Survey

Answer If Does your course instructor offer extra credit for participating in this study? Yes Is Selected

Q41 Please enter your email that you would like to receive the amazon gift card. Note: This does not need to be the same email that you received the email from.

APPENDIX D

DataBasics Matrix

Place an X under which DataBasic(s) you believe the item fits.

Item	BataBasic 1	DataBasic 2	DataBasic 3	DataBasic 4	DataBasic 5	DataBasic 6
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						

DataBasics:
(1) knowledge of diverse data collection protocols;
(2) selecting protocols best suited to answer educators' questions;
(3) collating and graphing data;
(4) discerning trends and differences in data;
(5) using data in team problem-solving; and
(6) selecting evidence-based interventions.

APPENDIX E

Original Test Blueprint

Knowledge of Data Collection Protocol	**Using Protocols Best Suited to Answer Team's Questions**	**Organizing and Graphing Data**	**Finding Trends and Differences in Data Questions**	**Using Data in Team Problem Solving Questions**	**Selecting Evidence Based Interventions**
Q1 When is an intervention evidence based?					
					x
Q2 What is baseline data?					
			x		
Q3 Which of the following is an example of a strong goal statement?					
				x	
Q4 Which of the following is an observable behavior?					
x					
Q5 How would you translate a student's behavior ratings of Rarely, Sometimes, Often, or Almost Always into data that could be used in a graph?					
Q6 Which of the following best describes an effective progress monitoring strategy?					
x	x				
Q7 Which of the following is an example of good evidence of student's progress in math?					
	x				
Q8 In addition to class grades, how could you reliably measure the academic performance of students in reading?					
x					
Q9 What should you do before collecting information on a student?					
				x	
Q10 What is an important consideration when choosing an intervention?					
x					
Q11 You have been monitoring the number of times a student was out of seat during class. Your data show that the student has good days and bad days but it is hard to tell if the student is improving. How could a graph show the parents whether the student is making progress?					
		x			
Q12 Pat is constantly disrupting class by being out of his seat. How would you measure how much Pat was out of his seat during an observation?					
x					

<table>
<tr><td colspan="6">Q13 Your team has collected data on a student with behavioral disorders for several weeks and is now ready to implement an intervention. How would you show where an intervention started on a line graph?</td></tr>
<tr><td></td><td></td><td>x</td><td></td><td></td><td>x</td></tr>
<tr><td colspan="6">Q14 A teaching team is worried about a student who is not passing her English class. The team’s data show that the student increased her work completion from 45% of assignments to 60% of assignments. Still, the student was failing because her grades on each assignment were still low. Given what the team knows, what would be their next step?</td></tr>
<tr><td></td><td></td><td>x</td><td>x</td><td></td><td></td></tr>
<tr><td colspan="6">Q15 A third grade teacher surveyed the students to see which subject was their favorite: math, science, reading, or social studies. If the teacher wanted to show the student’s a graph describing the percent of students preferring each subject, which graph should they choose?</td></tr>
<tr><td></td><td>x</td><td>x</td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Q16 A sixth grade teacher notices that her students appear to be arguing and complaining more than usual. She would like to collect data about what is taking place in her classroom. What kind of data collection would be useful in collecting the information she wants to measure?</td></tr>
<tr><td>x</td><td></td><td></td><td></td><td>x</td><td></td></tr>
<tr><td colspan="6">Q17 After collecting data on a student’s distracting behavior in class, a teacher wants to graph the number of times the student engaged in distracting behavior over a two week period on the line graph below. What would you label the x and y axes in the below graph?</td></tr>
<tr><td></td><td>x</td><td>x</td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Q18 A special education team met to make a plan for a student with a behavior problem. They defined the target behavior and created a rating scale from 0 to 4; with 0 representing a bad day and 4 representing a good day. They set a goal of the student earning a 3 or better each day. They collected data, but when they graphed it they were disappointed by how far the student was from meeting the goal they set. What step did the team forget in creating their plan?</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>x</td></tr>
<tr><td colspan="6">Q19 How could you describe the difference in data between the before intervention and during intervention phases on the graph below</td></tr>
<tr><td></td><td></td><td>x</td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Q20 A student with cognitive disabilities is being taught the steps of washing her hands. She routinely skips one or two steps and becomes frustrated. To figure out which steps of hand washing the students was skipping, what should the teacher do next?</td></tr>
<tr><td></td><td></td><td></td><td>x</td><td>x</td><td></td></tr>
<tr><td colspan="6">Q21 Your team was referred a student who has a history of being extremely difficult and resistant with frequent tantrums. Your team is not sure why the problem behavior occurs. What should they do?</td></tr>
<tr><td></td><td></td><td></td><td></td><td>x</td><td>x</td></tr>
</table>

Q22 A teacher has been implementing an academic intervention to increase a student's test scores. How should you describe the difference in data between the before intervention and during intervention phases on the graph below?					
		x			
Q23 You collected 5 days of baseline data and 5 more days of data after beginning an intervention. You graphed the data on a line graph but are unable to tell whether the student's on-task behavior improved. What could you do to clarify changes in the student's on-task behavior over time?					
		x	x		
Q24 A team of four teachers met briefly to create a plan for collecting data on a student who was consistently disruptive during their classes. They all decided to record the number of times the student was off-task during their class periods by tallying the number of times the student was disruptive. When they met after school, their tallies varied greatly and they could not agree on what the student's problem behavior was. What did they forget to do before collecting data on the student?					
		x	x	x	
Q25 You have been collecting data on a student for several weeks and decided to implement an intervention with the goal that your student would increase work completion from 45 percent to 80 percent of assignments. Your student has not missed completing an assignment for the last three weeks and appears to have reached this goal. What should you do next?					
			x	x	
Q26 A middle school team has been implementing an evidence-based intervention for student bullying. They are concerned that the teasing program might not be appropriate because 90% of their school's enrollment is Latino/a and the program was written for a low-income, predominately white college town in Southern Mississippi. Parents requested the school use an evidence-based program, but a counselor urges the team to use a peacemaking program that was written by a former counselor in the district because it is a better match to the students' culture. What should the team do?					
x				x	
Q27 How would you describe the trend test scores in the before and during phases of the graph below?					
		x	x		
Q28 An Art teacher wanted to record how much she was praising a First Grade classroom for positive behavior during a lesson. How could the teacher measure how much she praised the students?					
x			x		
Q29 Midway through the year, a newly-enrolled third grade student is referred to your student assistance team because of an inability read at grade level. How could your team gain the most useful information about this student's current reading abilities and instructional needs?					
x				x	

Q30 You are working with a first year teacher to use a behavioral intervention program to reduce classroom interruptions and teach cooperative work behaviors. Your role is to coach the teacher in using the intervention and provide the materials to implement it properly. After two weeks the teacher stopped the intervention because it was not working. What should you do next?					
	x				x

APPENDIX F

Comfort with Data Use Scale

Q1 I know a variety of reliable and valid methods for collecting data to describe a student’s academic and behavioral success

- 0 (0)
- 1 (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- 7 (7)
- 8 (8)
- 9 (9)
- 10 (10)

Q3 I use data daily to problem solve, monitor student progress, and make educational decisions about my students.

- 0 (0)
- 1 (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- 7 (7)
- 8 (8)
- 9 (9)
- 10 (10)

Q4 I am able to graph the data that I collect and use the graphs to recognize important trends and changes in student performance.

- ❍ 0 (0)
- ❍ 1 (1)
- ❍ 2 (2)
- ❍ 3 (3)
- ❍ 4 (4)
- ❍ 5 (5)
- ❍ 6 (6)
- ❍ 7 (7)
- ❍ 8 (8)
- ❍ 9 (9)
- ❍ 10 (10)

Q5 I feel comfortable sharing data with parents or teachers.

- ❍ 0 (0)
- ❍ 1 (1)
- ❍ 2 (2)
- ❍ 3 (3)
- ❍ 4 (4)
- ❍ 5 (5)
- ❍ 6 (6)
- ❍ 7 (7)
- ❍ 8 (8)
- ❍ 9 (9)
- ❍ 10 (10)

APPENDIX G

Subject matter expert consensus ratings and revised Blueprint

Knowledge of Data Collection Protocol	Using Protocols Best Suited to Answer Team's Questions	Organizing and Graphing Data	Finding Trends and Differences in Data Questions	Using data to refine instructional modifications	Selecting Evidence Based Interventions
Q1 When is an intervention evidence based? ❍ **a. When other people have used it and found that it worked well (1)** ❍ **b. When someone has collected pre and post data that shows meaningful change (3)** ❍ **c. When the author says that the intervention shows meaningful change (2)** ❍ **d. When it is published by a respected professional organization (4)**					
					X
Q2 How is baseline data different from intervention data? ❍ **a. Data that is collected over a short period of time (1)** ❍ **b. Data that is collected after an intervention is implemented (2)** ❍ **c. Data collected before an intervention has been implemented (3)** ❍ **d. The bottom 25 percent of all data points (4)**					
				X	
Q3 Which of the following is an example of a strong goal statement? ❍ **a. The student's on-task behavior will improve by the end of the semester according to teacher report. (2)** ❍ **b. The student's fluency will improve significantly over the semester as measured by a curriculum based assessment (3)** ❍ **c. The student's grades will be a "B" or better by the end of the semester according to report cards (4)** ❍ **d. The student will increase math computation scores by 15 points by the end of the semester as measured by curriculum based assessments (1)**					
				X	
Q4 Which of the following is an observable behavior? ❍ **a. The number of times a student blurts out in class (2)** ❍ **b. A student's attitude towards math or science class (1)** ❍ **c. The intensity of a student's feelings about a poor grade (3)** ❍ **d. The number of times a student becomes frustrated (4)**					
X					
Q5 How would you translate a student's behavior ratings of Rarely, Sometimes, Often, or Almost Always into data that could be used in a graph? ❍ **a. By counting the number of times each rating is used to describe the student's behavior (1)**					

<table>
<tr><td colspan="6">❍ b. By comparing the first day’s rating to the most recent day’s rating (2)
❍ c. By assigning a numerical value to each behavior rating (3)
❍ d. By grouping together the positive ratings in one graph and the negative ratings into a second graph (4)</td></tr>
<tr><td></td><td></td><td>x</td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Q6 Which of the following best describes a strategy to monitor a student’s academic progress?
❍ a. Weekly average scores for 20-word quizzes in English class (5)
❍ b. The daily number of words read correctly on selected reading passages (7)
❍ c. Total scores on a student’s mid-term and final in Algebra (6)
❍ d. Standardized test scores of a student’s cognitive abilities (8)</td></tr>
<tr><td></td><td>x</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Q7 Your kindergarten team is working with a student that is struggling with pre-literacy skills. What would you do to help the student succeed?
a. Look online for a reading intervention
b. Instruct the student at a different reading level
c. Change the student’s seat assignment
d. Retain the student for an additional year
e. Let the Principal decide</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>x</td></tr>
<tr><td colspan="6">Q8 In addition to class grades, how could you reliably measure the academic performance of students in reading?
❍ a. By asking the students how they are doing (2)
❍ b. By reviewing curriculum based measures benchmark scores (1)
❍ c. By reviewing their academic grades from the last two year (3)
❍ d. By giving the students a pop-quiz in one of their classes (4)</td></tr>
<tr><td>x</td><td>x</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Q9 What should you do before collecting information on a student?
❍ a. Collect baseline data on the student in his/her core classes (1)
❍ b. Decide how you are going to graph the data that you collect (2)
❍ c. Define and describe the identified problem (3)
❍ d. Meet with the student to talk about the identified problem (4)</td></tr>
<tr><td>x</td><td>x</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Q10 What is an important consideration when choosing an intervention?

❍ a. How easily is the intervention implemented (1)
❍ b. How will the data look once it is graphed (2)
❍ c. How popular is the intervention package in my district (3)
❍ d. How many students does the intervention affect (4)</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>x</td></tr>
<tr><td colspan="6">Q11 You have been monitoring the number of times a student was out of seat during class. Your data show that the student has good days and bad days but it is hard to tell if the student is improving. How could a graph show the parents whether the student is making progress?</td></tr>
</table>

- **a. Create a line graph that compares the student's out-of-seat behavior to that of a typical classmate (1)**
- **b. Graph the data and add a line that separates the before intervention and during intervention data (3)**
- **c. Collapse the data into weekly averages and graph them to decide whether these decrease over time (4)**
- **d. Graph the data and add a trend line that represents the trend in the student's out of seat behavior (2)**

		x	x		

Q12 Pat is constantly disrupting class by being out of his seat. How would you measure how much Pat was out of his seat during an observation?

- **a. Putting a hash mark on a piece of paper every time he left his seat (2)**
- **b. Starting a timer every time he leaves his seat and stopping it when he returns (1)**
- **c. Rating the disruptiveness of his out of seat behavior on a scale from 1 to 10 (3)**
- **d. Counting the number of students who stopped working when he left his seat (4)**

	x				

Q13 Your team has collected data on a student with behavioral disorders for several weeks and is now ready to implement an intervention. How would you show where an intervention started on a line graph?

- **a. Label where the intervention began on the data file worksheet (1)**
- **b. Create a separate graph for the pre-intervention and post intervention data (2)**
- **c. Draw a vertical line on the graph that separates the baseline data from the intervention data (3)**
- **d. Draw a line that shows the trend across pre and post intervention data (4)**

		x			

Q14 A teaching team is worried about a student who is not passing her English class. The team's data show that the student increased her work completion from 45% of assignments to 60% of assignments. Still, the student was failing because her grades on each assignment were still low. Given what the team knows, what would be their next step?

- **a. Gather data on work accuracy in addition to work completion (4)**
- **b. Shorten the length of the student's assignments (1)**
- **c. Place the student on a behavior plan (2)**
- **d. Gather data on the student's attention to work (3)**

				x	

Q15 A third grade teacher surveyed the students to see which subject was their favorite: math, science, reading, or social studies. If the teacher wanted to show the student's a graph describing the percent of students preferring each subject, which graph should they choose?

- **a. Stacked bar graph (1)**
- **b. Pie Graph (4)**

- ❍ **c. Line graph (2)**
- ❍ **d. Scatter plot (3)**

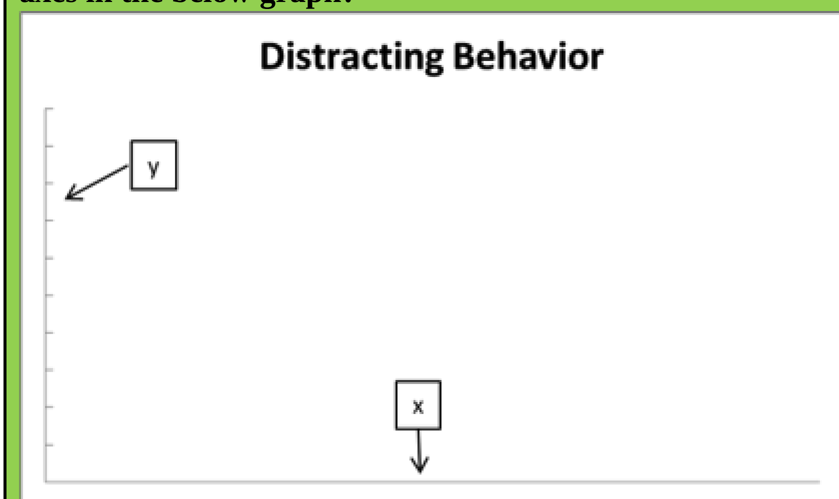
		x			

Q16 A sixth grade teacher notices that her students appear to be arguing and complaining more than usual. She would like to collect data about what is taking place in her classroom. What kind of data collection would be useful in collecting the information she wants to measure?

- ❍ **a. Tally the number of times students argued or complained in her classroom (2)**
- ❍ **b. Give the students a survey asking which subjects are their favorite at school (3)**
- ❍ **c. Consult fellow teachers for their thoughts and ideas on ways to improve the classroom climate (4)**
- ❍ **d. Collect anonymous survey data on the students' perceptions of the classroom environment (1)**

	x				

Q17 After collecting data on a student's distracting behavior in class, a teacher wants to graph the number of times the student engaged in distracting behavior over a two week period on the line graph below. What would you label the x and y axes in the below graph?



- ❍ **a. The number of behaviors on the x-axis and the class periods on the y-axis (1)**
- ❍ **b. The date on the x-axis and the number of distracting behaviors on the y-axis (2)**
- ❍ **c. The weekly averages on the x-axis and the number of distracting behaviors on the y-axis (3)**
- ❍ **d. The date on the x-axis and the number of times the student left his seat on the y-axis (4)**

		x			

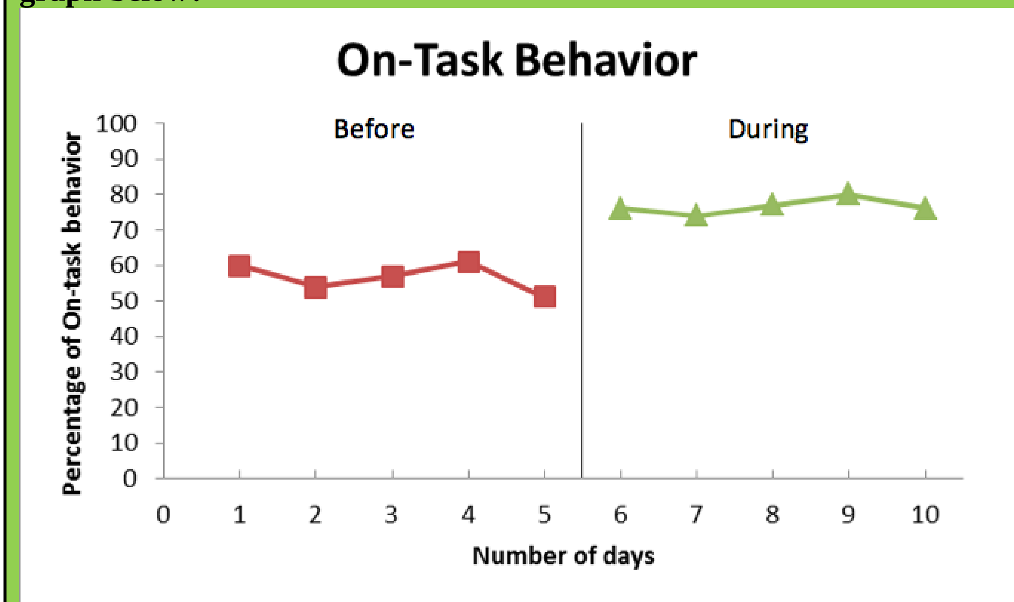
Q18 A special education team met to make a plan for a student with a behavior problem. They defined the target behavior and created a rating scale from 0 to 4; with 0 representing a bad day and 4 representing a good day. They set a goal of

the student earning a 3 or better each day. They collected data, but when they graphed it they were disappointed by how far the student was from meeting the goal they set. What step did the team forget in creating their plan?

- ❍ a. Collect data describing the student's initial behavior (1)
- ❍ b. Rate the student's behavior in multiple settings (3)
- ❍ c. Define a goal related to the student's behavior (2)
- ❍ d. Record the student's behavior in a continuous manner (4)

				x	

Q19 What accounts for the differences between the before and after phases on the graph below?



- ❍ a. It is impossible to tell because there was too much variability in each phase (1)
- ❍ b. The intervention was not effective because the student became less on-task after the intervention was implemented (2)
- ❍ c. The intervention was effective in increasing the student's on-task behavior because the data in the during phase do not overlap with data in the before phase (3)
- ❍ d. There are not enough data points to determine if the intervention had an effect (4)

			x	x	

Q20 A student with cognitive disabilities is being taught the steps of washing her hands. She routinely skips one or two steps and becomes frustrated. To figure out which steps of hand washing the students was skipping, what should the teacher do next?

- ❍ a. Count every time the student leaves to wash her hands and record it at the end of the day (2)
- ❍ b. Put a tally mark on a piece of paper every time the student washes her hands correctly (3)

- c. Rate how well the student washed her hands on a scale from 1 to 10 and graph her daily percentages (4)
- d. Write the hand washing steps in order and circle which ones she completes (1)

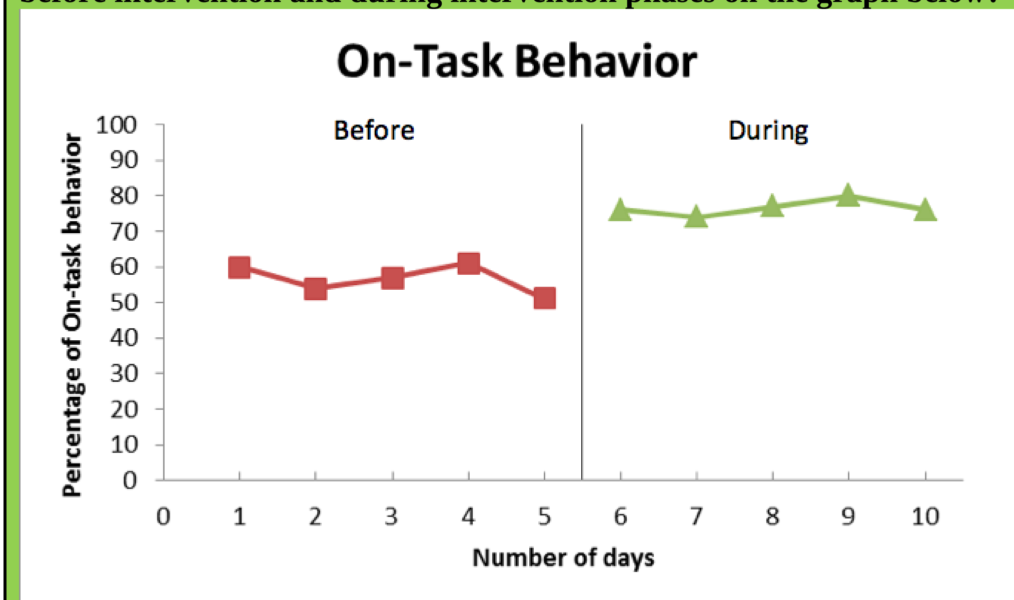
	x		x		

Q21 Your team was referred a student who has a history of being extremely difficult and resistant with frequent tantrums. Your team is not sure why the problem behavior occurs. What should they do?

- a. Use a broad screening measure to gather data on behaviors, academic skills, participation, and learning for the student (4)
- b. Implement an evidence based intervention program that reduces the tantruming behavior (1)
- c. Call an IEP meeting to determine whether to qualify the student for special education services (2)
- d. Observe the student at lunch and recess and tally the number of times the student tantrums (3)

	x				

Q22 A teacher has been implementing an academic intervention to increase a student's test scores. How should you describe the difference in data between the before intervention and during intervention phases on the graph below?



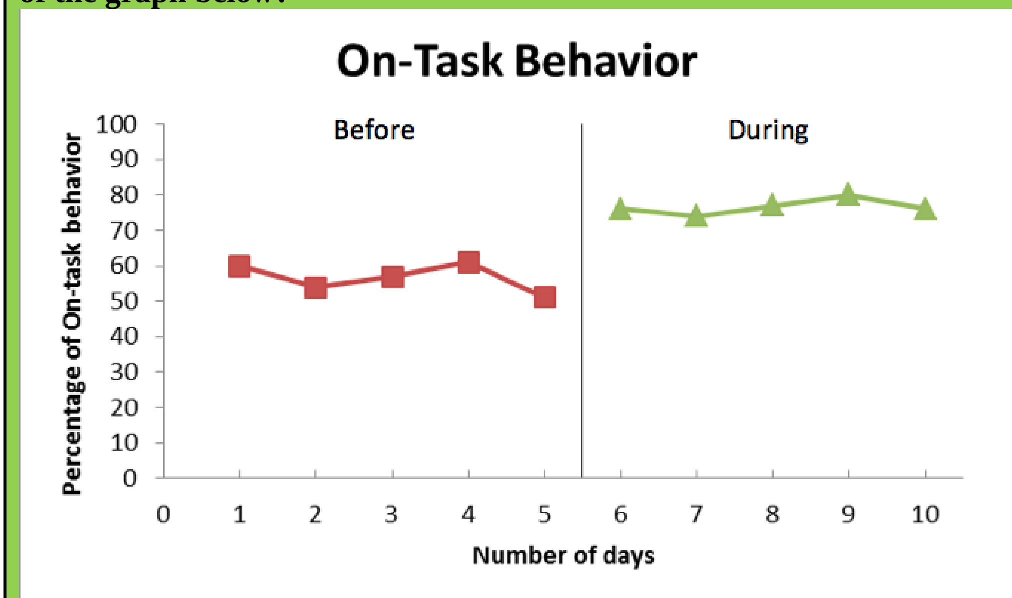
- a. The intervention was effective in increasing the student's quiz scores because data in the during phase do not overlap with data in the before phase (2)
- b. The intervention was not effective in increasing the student's quiz scores because there was not a significant difference across phases (1)
- c. The intervention was not effective because the student's quiz score decreased after the intervention was implemented (3)
- d. There are not enough data points in before and during phases to determine if the intervention had an effect (4)

<table>
<tr><td></td><td></td><td></td><td>x</td><td></td><td></td></tr>
<tr><td colspan="6">Q23 You collected 5 days of baseline data and 5 more days of data after beginning an intervention. You graphed the data on a line graph but are unable to tell whether the student’s on-task behavior improved. What could you do to clarify changes in the student’s on-task behavior over time?
❍ a. Draw a horizontal line on the graph to show your goal for the student’s behavior (1)
❍ b. Draw a trend line on the graph to show the 10 day pattern of the student’s on-task behavior (3)
❍ c. Separate baseline and intervention data and draw a trend line for each phase (2)
❍ d. Change the graph type to bar graph (4)</td></tr>
<tr><td></td><td></td><td>x</td><td>x</td><td></td><td></td></tr>
<tr><td colspan="6">Q24 A team of four teachers met briefly to create a plan for collecting data on a student who was consistently disruptive during their classes. They all decided to record the number of times the student was off-task during their class periods by tallying the number of times the student was disruptive. When they met after school, their tallies varied greatly and they could not agree on what the student’s problem behavior was. What did they forget to do before collecting data on the student?
❍ a. Define the target behavior in precise terms (1)
❍ b. Observe the student during recess (2)
❍ c. Have frequent meetings about the student (3)
❍ d. Decide how often they would tally the behaviors (4)</td></tr>
<tr><td></td><td>x</td><td></td><td></td><td>x</td><td></td></tr>
<tr><td colspan="6">Q25 You have been collecting data on a student for several weeks and decided to implement an intervention with the goal that your student would increase work completion from 45 percent to 80 percent of assignments. Your student has not missed completing an assignment for the last three weeks and appears to have reached this goal. What should you do next?
❍ a. Continue the intervention and data collection (1)
❍ b. Discontinue collecting and recording the student’s progress (2)
❍ c. Implement an additional interventions that supports work completion (3)
❍ d. Decrease the intervention and continue to monitor the student’s work completion for two more weeks (4)</td></tr>
<tr><td></td><td></td><td></td><td></td><td>x</td><td></td></tr>
<tr><td colspan="6">Q26 A middle school team has been implementing an evidence-based intervention for student bullying. They are concerned that the teasing program might not be appropriate because 90% of their school’s enrollment is Latino/a and the program was written for a low-income, predominately white college town in Southern Mississippi. Parents requested the school use an evidence-based program, but a counselor urges the team to use a peacemaking program that was written by a former counselor in the district because it is a better match to the students’ culture. What should the team do?</td></tr>
</table>

- a. They should adapt the evidence-based program to fit the culture of their students (4)
- b. They should choose the evidenced based intervention because it has been empirically tested (1)
- c. They should choose the peacemaking program because it was designed for the culture of the district (2)
- d. They should drop the peacemaking program because it is not evidenced based (3)

				x	x

Q27 How would you describe the trend test scores in the before and during phases of the graph below?



- a. The before trend is increasing and the during trend is decreasing (1)
- b. Both the before and during trend are increasing (2)
- c. The before trend is decreasing and the during trend is increasing (3)
- d. The before trend is improving and the during trend is worsening (4)

			x		

Q28 An Art teacher wanted to record how much she was praising a First Grade classroom for positive behavior during a lesson. How could the teacher measure how much she praised the students?

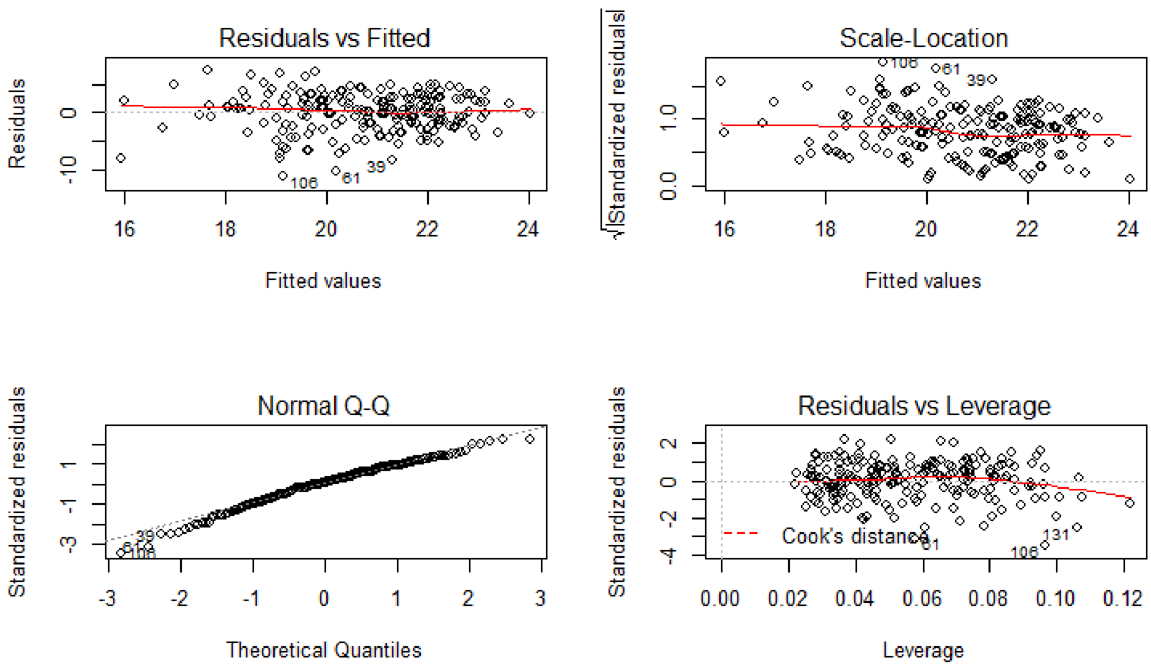
- a. Starting a timer every time a student was praised and stopping it when the teacher had to redirect a student (2)
- b. Putting a hash mark on a piece of paper every time the teacher praised students and adding the total praises at the end of the lesson (1)
- c. Rating how well the teacher praised students on a scale from 1 to 10 and graphing the results (3)
- d. Counting the number of students the teacher had a positive interaction with and recording the total (4)

	x				

<table>
<tr><td colspan="6">Q29 Midway through the year, a newly-enrolled third grade student is referred to your student assistance team because of an inability read at grade level. How could your team gain the most useful information about this student’s current reading abilities and instructional needs?
❍ a. Conduct observations of the student during reading instruction (1)
❍ b. Give the student a standardized achievement battery and cognitive abilities test (2)
❍ c. Collect information about the student’s quarterly grades in all academic areas (4)
❍ d. Use curriculum based measures to track the student’s progress in reading (3)</td></tr>
<tr><td>x</td><td>x</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Q30 You are working with a first year teacher to use a behavioral intervention program to reduce classroom interruptions and teach cooperative work behaviors. Your role is to coach the teacher in using the intervention and provide the materials to implement it properly. After two weeks the teacher stopped the intervention because it was not working. What should you do next?
❍ a. Observe the teacher during a class to see if the intervention was implemented as it was written (1)
❍ b. Discontinue the intervention and try another evidence-based program (2)
❍ c. Refer the problem students for special educational services because they are resistant to intervention (3)
❍ d. Add additional behavioral interventions to the classroom to see if these improve classroom behavior (4)</td></tr>
<tr><td></td><td></td><td></td><td></td><td>x</td><td>x</td></tr>
</table>

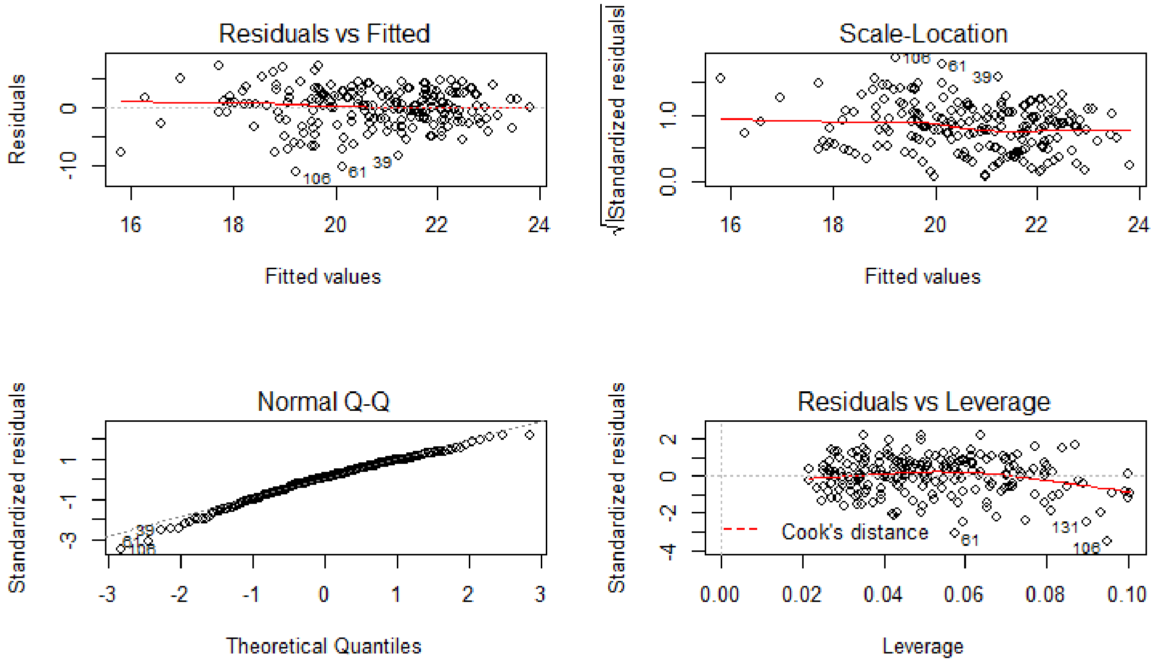
APPENDIX H

Plots of model 1020



APPENDIX I

Plots for model 923



APPENDIX J

Cooks Distance for Model 923