SMU | ANNETTE CALDWELL SIMMONS SCHOOL OF EDUCATION & HUMAN DEVELOPMENT

RESEARCH IN MATHEMATICS EDUCATION

Teacher Advisory Panel Technical Report

Fall 2018-Summer 2019

Technical Report 19-25

Teacher Advisory Panel Technical Report Fall 2018-Summer 2019

Tina Barton• Kindra Knight • Cassandra Hatfield • Lindsey Perry• Leanne Ketterlin Geller

Southern Methodist University

Summer 2019

Published by
Southern Methodist University
Department of Education Policy & Leadership
Simmons School of Education & Human Development
PO Box 750114
Dallas, TX 75275-0114
Contact information: rme@smu.edu

This research was supported by The National Science Foundation (NSF); 1721100. Opinions expressed herein do not necessarily reflect those of NSF or individuals within.

Acknowledgements: We would like to thank Dr. Jennifer McMurrer, Alain Mota, Erica Simon and Robyn Pinilla for their assistance on this publication, including reviewing and editing the documents.

Executive Summary

The Measuring Early Mathematical Reasoning Skills (MMaRS) project includes an important feedback component that creates opportunities throughout the development process to incorporate perspectives and input received directly from classroom teachers and other potential users who will engage with the tools that MMaRS is working to develop. The Teacher Advisory Panel (TAP) was included to support the work that the MMaRS project is doing to create assessment instruments and resources for K-2 educators. The TAP provides a unique window into the actual practice of K-2 educators that can enrich the research that MMaRS is conducting by employing Human-Centered Design (HCD) research methods and principles to ensure that the project goals and resulting deliverables are as effective as possible, based on direct user feedback. The purpose of this technical report is to explain the research activities and findings from engagement with the TAP during the 2018-19 school year.

Human Centered Design is a methodology that bridges the implementation of design and research to engage end-users of a product or system at each phase of the development process. According to Giacomin (2014), HCD is "based on the use of techniques which communicate, interact, empathize and stimulate the people involved, obtaining an understanding of their needs, desires and experiences which often transcends that which the people themselves actually realised." (p. 609)

By utilizing principles of HCD in focus group settings during this phase of development, MMaRS has garnered insights and perspectives from practitioners related to their needs and experiences which will inform design decisions throughout the project. As a result of the TAP data collected MMaRS will be able to specifically address opportunities for design that will impact the usability and usefulness of the assessment instruments developed. Ultimately, this will result in an increased likelihood of buy-in and long-term implementation by practitioners in the field, positively impacting student learning and assessment experiences.

Table of Contents

Teacher Advisory Panel Technical Report: Fall 2018-Summer 2019

Introduction

Measuring Early Mathematical Reasoning Skills (MMaRS) has as its goal to create assessment instruments to measure numeric relational reasoning and spatial reasoning with K-2 students in classroom settings. The data acquired from these assessment instruments will inform instructional decisions that teachers make to improve student learning related to these constructs. The MMaRS project includes a variety of important components and processes that work together to inform the development and design of these assessment instruments.

The Teacher Advisory Panel (TAP) is a vital component of the MMaRS project. The purpose of this panel is to better understand how they use data in their classrooms, what types of decisions they want to make using data and what makes data useful and usable. To this end, the TAP project included traditional and interdisciplinary methods of data collection. Teachers were invited to participate based on recommendations from Dallas-area public and charter school district administrators and educational colleagues, giving the data depth of experience and reflecting a high-caliber of expertise in the field. The implications of this research and the potential of this work to positively impact assessment development and the practitioners who utilize these tools is far reaching. The inclusion of Human-Centered Design (HCD) methods in this component of the project and the innovative approach to this work has produced relevant, meaningful data that represents the values and perspectives of practitioners in K-2 education. The insights garnered from these data can inform next steps in both prototyping and future design opportunities in assessment development for the improvement of user experience.

Context

Teachers and other educational support professionals work in a fast-paced, dynamic, and cognitively demanding environment. The degree of analytical consideration and emotional engagement required to effectively manage a classroom of 20-30 young human beings is a feat in and of itself. Additionally, educators calculate the consistent fluctuation of resources and changing needs of learners, day by day, adapting and adjusting instructional plans and decisions to meet the needs of their learners as best they can in the classroom setting. Teachers' perspectives are invaluable to us as assessment researchers, educational designers, and public policy-makers (Weingarten, 2012). Grappling with the changing needs of our students to adapt and develop in order to meet the ever-changing demands of a yet-to-be-defined future of work, coupled with the challenges related to equity and STEM education in the current educational landscape, the goals of the TAP project scratch only the surface of how we can aspire to engage with the greatest resources we have in our work to redefine the system of education in order to best prepare our learners for their future work and lives in the global economy and society to come.

Factors

Assessment is a big word in education. It can mean many things to any given teacher: It is a tool for quantifying the qualitative data teachers have collected about their learners. It is a measure of teachers' instructional practices and their learners' abilities to receive information conveyed throughout the year. It provides a way for students to demonstrate their learning. To be effective, Glaser, Chudowski, and Pellegrino (2001) emphasize classroom assessment "should focus on making students' thinking visible to both their teachers and themselves so that instructional strategies can be selected to support an appropriate course for future learning" (p. 4). This form of assessment gives us both a window to see into our learners' growth and development as well as a pathway to access how to better help learners into the next stage of their academic journey.

Based on findings from the TAP meetings, teachers have mixed feelings about assessment and its resulting data. School districts struggle to make assessment data relevant to practice and to facilitate the clear communication of student growth and development over the course of their educational journey within the district.

Participants

Teachers

The MMaRS research team elicited recommendations from local public and charter school districts and educational colleagues within the Dallas-Metroplex to compile a list of highly-qualified elementary educators that would be willing to participate in an ongoing Teacher Advisory Panel for the duration of the MMaRS project. To support nominators in identifying potential teachers to serve on the TAP, the following guidelines were included in the call for nominations:

- At least 3-5 years of experience in K-2 mathematics
- A commitment to integrating assessment data to inform instructional design
- Curiosity and enthusiasm for early mathematics and willingness to explore these reasoning concepts
- Interest in developing an ongoing relationship with SMU that integrates research and practice in meaningful ways

Eleven teachers were nominated to participate. The MMaRS research team identified that all of the nominated teachers fit the criteria and spanned across grade levels and years of experience. These eleven teachers were contacted via email and given information about the TAP and asked if they would like to participate. Of the eleven that were nominated, eight teachers accepted the nomination.

The initial request as a part of the TAP was to participate in focus groups at different times of the project, which would provide insights and experiences related to assessment theory and practice in their current classroom settings. This information would be used to inform design decisions and would provide a source of feedback during prototyping and the next steps of the MMaRS project. Teachers were offered compensation for each focus group meeting that they participated in and were fed a meal prior to each meeting.

As depicted in figure 1, the MMaRS TAP included eight K-2 teachers with varying degrees of educational experience, years of service, and teacher-leadership roles within their districts. The districts that TAP teachers represent are varying size public-school districts and one charter district from around the Dallas-Metroplex.

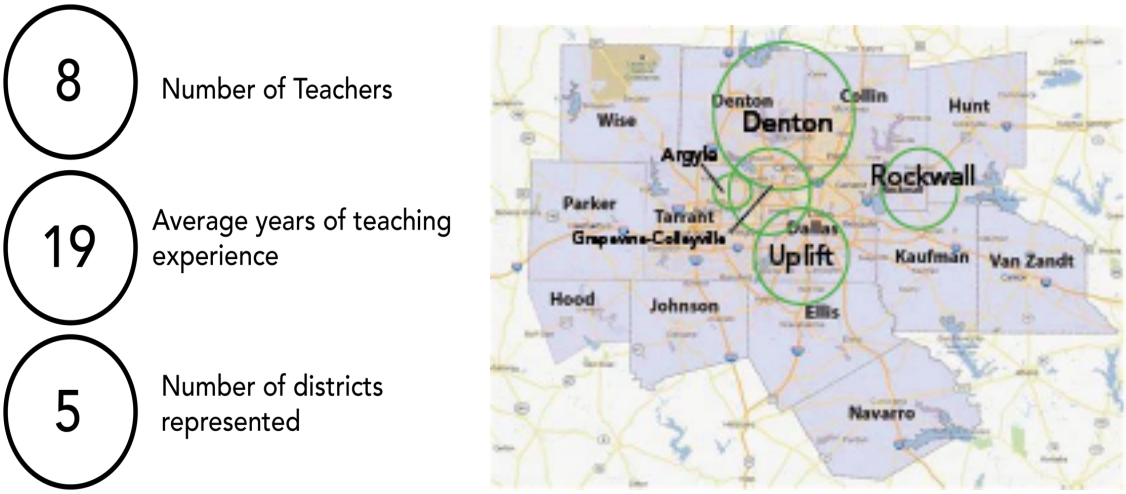

Figure 1. TAP experience and district

Learners

The frame of reference that our TAP has been asked to consider are the children they currently teach with in their K-2 general education classroom settings. This includes students with disabilities, students from a variety of sociocultural and socioeconomic backgrounds, as well as students with different home environments.

The TAP includes teachers from a varied representation of public-school districts from around the Dallas Metroplex. These districts include a demographic make-up of primarily Caucasian students (51% or higher each district) with the exception of a public charter school network which has a student population that includes 65% Hispanic students.

Administrators

The TAP teachers operate within school settings that have a traditional staff of school administration and district affiliation. Again, the exception in this sample is a teacher who teaches in a public charter school system. This model includes different mandates for

assessment and practices for reporting that are captured in the data from both the fall and the spring meetings.

Families and the community

The TAP teachers represent a variety of different learning communities and school districts from around the Dallas Metroplex. Some teachers are practicing in urban communities, while others teach in more suburban settings. The demographic make-up of each of these communities vary based on the surrounding areas and the families of the children served in our TAP teachers' classrooms.

Research Questions

Circling back to the goal of the MMaRS project to create assessment instruments to measure numeric relational and spatial reasoning with K-2 students in classroom settings, we posed the primary question: How might we develop an assessment and corresponding learning progression that is valuable to teachers in the classroom? The purpose of the TAP research outlined in this paper is to inform this MMaRS project goal.

More specifically, the research questions we considered are similar to the design process "How might we?" questions that seek to understand a particular context and the role of a system or tool within that context. (See Figure 2 for an illustration of the useful and usable concepts.) These included the following:

- What are key areas of application to prioritize when designing tools to make data useable for practitioners?
    - How do teachers use data in their communication about assessments?
    - What makes data useful in instructional design?
    - What areas of instructional design do teachers use data for?
    - What are the struggles that classroom teachers encounter in using data?
    - What are the wins that classroom teachers encounter in using data?
- What factors influence teachers' use of data in the classroom?
    - How do teachers want to use data?
    - What role do data play in teachers' day-to-day decision making?
    - What makes data useful to classroom teachers?
    - What makes data useable to classroom teachers?

Because the TAP research being conducted is a process focused specifically on the users involved in the system of assessment and development, it was fitting for our main research methods to be based in both traditional qualitative research and human-centered design methodologies.

Valuable is defined as both useful and useable in this context

Useful is effective for accomplishing a task.

Useable is effective, and enjoyable or pleasant.

For example, a traditional can opener when originally designed was very useful for accomplishing the task of opening cans that were inaccessible (without blunt force and sharp objects) otherwise. It was not, however, very comfortable to use nor did it look particularly fun, attractive, or inviting.

When the can opener was redesigned to address the ergonomic and aesthetic considerations that consumers had been experiencing for years, the new can openers literally "opened the way" for many more versions of both *useful* and *useable* versions to be developed. Now you can have a can opener that not only opens cans, but also matches your kitchen décor and doesn't cause your hand to cramp or blister while opening that second can of sauce.

Figure 2: Useful Useable Definitions

Method

Human-Centered Design

Human-Centered Design (HCD) is a methodology that bridges the implementation of design and research to engage end-users of a product or system at each phase of the development process. According to Giacomin (2014), HCD " is based on the use of techniques which communicate, interact, empathize and stimulate the people involved, obtaining an understanding of their needs, desires and experiences which often transcends that which the people themselves actually realised" (p. 609).

By utilizing principles of HCD in focus group settings during the initial phase of development, MMaRS has garnered insights and perspectives from practitioners related to their needs and experiences which will inform design decisions throughout the project. As a result of the TAP data collected, MMaRS will be able to specifically address opportunities for design that will impact the usability and usefulness of the assessment instruments developed. Ultimately, this

will result in an increased likelihood of buy-in and long-term implementation by practitioners in the field, positively impacting student learning and assessment experiences.

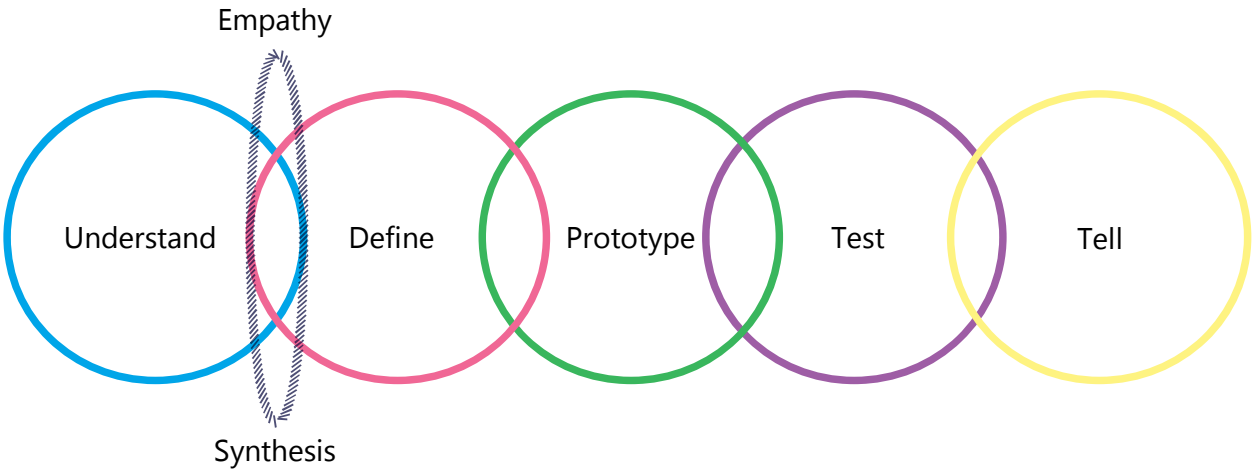

Figure 3: Human-Centered Design Process

Human-Centered Design is defined as the application of a strategic research methodology to better understand a given context with the goal of improving a process, product, or experience. Faculty within the Master of Arts in Design and Innovation (MADI) program at SMU define HCD as a creative approach to problem-solving that designs with the end-user in mind. Drawing from the MADI course description, HCD is a "well-established process and set of methods aimed at devising solutions based on people's needs" (SMU 2020). It is a methodology employed by different groups and fields and most companies generate their own process that they follow. (J. Burnham, personal communication, May 7, 2020). The process taught by co-founders of the MADI program at SMU, Kate Canales and Gray Garmon, was adapted for MMaRS researchers' inquiry with the TAP, as shown in figure 3.

The process includes 5 general phases of implementation depicted in figure 3, defined as: Understand, Define, Prototype, Test, and Tell. Within the Understand phase of the HCD methodology, researchers seek to *understand* deeply a given context by engaging with participants and environments to generate empathy. Secondary and primary research is conducted in order to collect data that can help *define* a design opportunity within the given context. By including primary sources of research such as interviews, focus group data, and observations; the end-users remain central to the process and the development of future prototypes and products are grounded in the actual users' perspectives and needs.

Once data is collected and externalized, researchers analyze data to identify themes within the data. These themes can then be synthesized to harvest insights related to the needs of the end-user and identify opportunities for *prototyping* to answer questions that still need to be addressed. Synthesis of the data and themes identified involves both the application of analytical thought and a developing holistic understanding of the needs of the users. By developing a

deeper understanding of the context and needs, designers can begin to develop prototypes that will either directly or indirectly address the users' given needs.

Once prototypes have been developed (often through rapid, low-resolution iteration), they are *tested* to find answers to further questions and iterated on to the solve for the greater design challenge. Once testing of prototypes is complete and findings are determined, solutions and/or possible recommendations can be made through the *tell* phase. In traditional research, this may be considered the dissemination phase of the design research process.

During the first year of the TAP project, researchers engaged in many elements of the Human-Centered Design process. These elements included the understand phase and the define phase, as depicted in the above graphic (see Fig. 3). Through a thorough externalization and synthesis of the year one data collected, HCD researchers were able to extrapolate insights and recommendations for further prototyping and focus areas for future ideation with regard to the MMaRS development of assessment instruments for Numeric Relational Reasoning and Spatial Reasoning.

Understand

In the Understand phase of the HCD process, researchers seek to understand the context and factors that influence a given design challenge. The MMaRS TAP research team is seeking to understand the mathematics assessment experience for classroom teachers. Specifically, the TAP consists of K-2 teachers who engage in a variety of assessment processes to identify student needs, groups students, and monitor progress towards district, state, and national standards for mathematics growth and development. The TAP practitioners participated in two focus groups, in both the Fall and Spring of the 2018-2019 school year. We provided opportunities for each practitioner to share their experience and goals around mathematics assessments. We also elicited their perspectives related to data usability and the relationship between assessment and instructional practices.

The most important opportunity within the HCD phase of understanding the design challenge is to develop empathy for the end-users of a given product or system (Bowie & Cassim, 2016; Hanington, 2003; IDEO.org, 2015). By seeking to understand deeply, the divergence of data and thought-processes related to the design challenge can provide ample opportunity for innovative thought and creative understanding that will inform the ideation phase (IDEO.or, 2015). With the end-users central to Understand and Define phases of the design process, we are empowered with authentic insights and perspective that inform our definition of the design challenge. This results in a process of co-creation in ideation that can produce powerful results.

Define

The TAP research team has sought to better understand the environment, needs, and constraints of the classroom teachers and their experience of mathematics assessment in its current state. Ultimately seeking to answer the question "how might we develop an assessment tool that is valuable to teachers?"

By utilizing both open-ended (Fall Focus Group and Affinity Diagramming) and more narrowly focused prompts (Love Letters and Break-up Letters), the data that MMaRS researchers collected from TAP has provided rich insights into the practical assessment experience of K-2 teachers, as well as their values related to assessment and data. We also gleaned a better understanding of the themes associated with teachers' varied experiences and are able to identify key insights that result from our analysis.

Identifying these themes and corresponding insights will inform our next steps and recommendations as we move into the ideation and prototyping phase of our research.

Data Sources

As previously outlined, the goal of the MMaRS project is to create assessment tools to measure numeric relational reasoning and spatial reasoning with K-2 students in classroom settings. These assessment instruments would ideally be useful and useable to practitioners in a general education setting. The data acquired from these assessment instruments will inform instructional decisions that teachers make to improve student learning related to these constructs. The correlation between an increased competency with numeric relational reasoning and spatial reasoning during the early elementary years and a higher level of achievement and performance in algebraic thinking and reasoning in high school and beyond is at the heart of this project. The MMaRS project includes a variety of important components and processes that work together to inform the development and design of these assessment instruments. The Teacher Advisory Panel (TAP) is one of those components. The TAP provides an avenue for MMaRS researchers to better understand the relationship that practitioners have with assessment and data in the classroom. A summary of the data sources is included in table 1 and with more detail about the data collection method and activities in the narrative that follows.

Table 1.

Summary of Data Sources

Method Used	Participants	Context Collected	Date Collected
Focus Group	TAP teachers (ALL- 8)	Fall TAP Meeting	October 25, 2018
Dots Activity	TAP Teachers (6)	Spring TAP Meeting	May 9, 2019
Love Letter/Break-up Letter	TAP Teachers (6)	Spring TAP Meeting	May 9, 2019
K.J. Technique	TAP Teachers (6)	Spring TAP Meeting	May 9, 2019

Focus Group Activities

Focus Groups are defined as "a small group of people whose response to something (such as a new product or a politician's image) is studied to determine the response that can be expected from a larger population" (Merriam-Webster, n.d.) The purpose of the TAP focus groups in the process of HCD was to better understand the context and users of assessment instruments in K-2 education.

The intentional design of each experience for focus group participants reflected the thoughtful consideration of the users' participation, affective engagement, and overall comfort level for sharing information within the context of the focus group. Meals were served at each focus group and chairs were arranged around an oblong table. There were microphones and laptops present, but the emphasis was on face-to-face participation and engagement with one another. Teachers were encouraged to share their thinking and care was taken to ensure that each participant had an opportunity to participate and contribute to the conversation.

Fall 2018: Moderated Discussion of classroom-based practices and philosophies- self-reporting

In the understand phase of HCD, we seek to understand users' context, including the tools and processes they use, how and why they interact with these tools and processes in the way that they do, and how they feel when doing so. As such, for this group of TAP teachers, we sought to understand what curriculum and assessments the teachers utilized in their classrooms, how and why they used them, and what teachers felt the result or impact was of said curriculum and assessments. TAP teachers were invited to participate in a focus group for two hours.. Eight teachers in total joined the MMaRS research team in a conference room on Southern Methodist University's campus. Teachers were gathered in the conference room around a large oblong table with Dr. Leanne Ketterlin-Geller at the far end and MMaRS researchers seated intermittently among the teachers. The moderated discussion was recorded using microphones that were visible to the participants and care was given to instruct participants to speak loudly in order to capture their thinking as completely and effectively as possible. Observational notes were also collected by researchers that were present in the room. Teachers were warmly welcomed with a meal and encouraged to connect with one another while waiting to begin the session's activities. By intentionally creating an inviting environment with a meal and built-in time to socialize, the atmosphere of collaboration and community were reinforced. Teachers introduced themselves and made connections with one another while learning about background experience and current placements.

The TAP teachers were prompted with a series of questions that Dr. Ketterlin-Geller moderated using questions that were posted on the screen for teachers to see. Once the focus group began, teachers began to express their thoughts and opinions related to experiences with data and assessments in both Reading and Mathematics. Teachers engaged in a lively conversation around instructional practices and limitations in both ELA and Mathematics in their classrooms and respective districts. Teachers shared experiences and opinions related to their use of data, score reports, and other tools for assessment and data collection. The participants verbalized responses to the prompts and gave examples from their actual practice. Discussions often ensued that provided deeper insights into the actual experience of data collection and assessment in the context of primary classrooms.

Spring 2019: Experience-based focus-group with participatory elements of HCD research methods and verbal/non-verbal prioritization

During the spring TAP meeting, MMaRS researchers engaged with the Define phase of the Human-Centered Design process, seeking to define a framework of values to move forward with into the prototyping stage. Researchers implemented a number of specific methods in order to

gather data to answer the research question "how we might make assessment data useful and useable to teachers?"

K-2 Teachers were invited to participate in another focus group on May 9th, 2019. Six teachers of the original eight were able to attend this meeting. The meeting lasted for two hours and took place in the conference room of the RME office. This meeting consisted of carefully designed experiences to elicit participants' thoughts and feelings. Human-centered design data collection methods were selected that would initiate participation by all and eventually cultivate and generate a consensus from the whole. These HCD methods included dot (heat mapping), love letter/break up letter, and affinity diagramming/KJ technique exercise. Each of these activities is outlined in more detail below.

Dot Exercise (heat mapping). The dot exercise, sometimes referred to as heat mapping, was a method we used to collect individual, quantifiable data related to specific statements and key insights that were extracted from the dialogue we engaged in during the fall focus group. The insights centralized around how data are used to make decisions for grouping and next steps in instruction and planning. This relates to several of the research questions such as "what role do data play in teachers' day-to-day decision making" and "how do teachers want to use data?"

Participants received six green and six red "dots" upon arrival. There were three prepared posters that had nine statements each, all ranging in content from grouping decisions to instructional data use. Teachers were then prompted to apply the dots according to their level of priority for each of the posters, using six dots per poster. Green dots indicated high priority and red dots indicated lowest priority. Participants received six new dots for each poster. Each poster had nine statements to consider. The combination of dots per poster reflected participants agreement or disagreement with the prioritization of each given statement. No dot indicated no prioritization or consideration. This activity was conducted in an open-ended way as an entry task while people continued to arrive, so participants were able to move at their individual pace from one poster to another and back again.

Love Letter/Break-up Letter (Affective expression of personal perspective). A design research strategy used to elicit participants hopes and dreams, as well as fears and frustrations; the Love Letter/Break-up Letter method helps designers "understand the less tangible aspects of the things they create; specifically, the social, human values and meanings conveyed through the things and experiences we design, as well as their understandability and usability" (Shedroff 2003, p. 159).

This method is a brand strategy that was modified to explore the complicated and nuanced relationship that teachers have with assessment data. The activity begins with a prompt designed to elicit emotional reactions (Martin & Hanington, 2012). A "love letter" and a "break-up letter" are typically written expressions of the emotions one feels with regard to an intimate relationship with a person, whether at the beginning/middle/end of that relationship. When implementing this method, care must be taken to account for the emotional nature of the prompt. People may have strong reactions, even unexpectedly, if triggered by emotional trauma etc. As a precaution, the facilitator of this activity engages in intentional relationship and community building interaction, connecting with common interests and finding neutral ground with the group as a whole. Genuine communication and humor are powerful tools for trust building and their importance cannot be overstated when employing HCD data collection methods. This

provides the opportunity for a level of safety amongst the participants that will encourage honest and open sharing. That sharing issues in a richer and fuller data source in the long run. The teachers were informed of the researchers' intent to collect the letters, however, they were also reassured that the letters would not be displayed publicly with any identifying information, further assuring emotional and social safety.

Teachers were prompted with a "Dear Assessment Data," letter format for both a "Love Letter" and a "Break Up Letter". An example of the format was shared verbally with unrelated content, so as not to influence the participants. The letters shared modeled what it sounded like to personify an inanimate object/system- namely, the postal system and another well-known delivery system. These examples were developed to encourage creativity and foster a level of comfort with the type of writing being requested. Teachers were provided with approximately 10 minutes to write the letters and given the opportunity to share out when finished. Teachers were allowed to choose which letter to start with and were given time to share out after everyone was done writing.

Affinity Diagramming/KJ Technique (Consensus Building). Affinity diagramming, also referred to as the KJ Technique was used with the TAP as a brainstorming and consensus building activity.

- Affinity diagramming:

  A collaborative method employed within the process to organize ideas, themes, and priorities by "creating a visual representation of a team's observations, knowledge, concerns, and ideas" (Martin & Hanington, 2012, p. 104).

- The KJ technique:

  A Human-Centered Design method used for consensus building that is silent, makes effective use of time, and negates "group pressure" by providing personal representation of perspective with minimal judgement passed within a group setting (Kawakita, 1982).

This method is employed in business and design industry meetings for the purpose of brainstorming valuable data and building consensus among large groups within a limited amount of time. The KJ technique, specifically, originated in Japan as a method employed during business meetings to work through difficult challenges in a large group while providing every participant with a voice. The method is traditionally completed silently until the participants are asked to discuss titles for each group that is determined (see Step 2 in Figure 4 Spring TAP Affinity Diagramming). Each participant has an opportunity to write their ideas on sticky notes and all are encouraged to write as many ideas down as possible, even building off of one another's ideas and spurring further brainstorming. Once ideas have been generated, there is a period of silent sorting that takes place where similar ideas are grouped together to identify themes.

Figure 4: Affinity Diagram/KJ Technique

Teachers were prompted with the question "What makes data useful and useable to you?" and were asked to write ideas on sticky notes and post them on the wall. A facilitator guided the group through the process of grouping and regrouping ideas silently, then through the step of identifying themes and creating names or titles for each group, first individually, then collectively. Teachers were given a different color sticky note for this step (blue). Teachers were prompted to be silent during the second step of the process, however more discussion was had around titles of groups and researchers determined this would be allowed for the purpose of collecting thoughts verbalized that were not explicitly written on sticky notes. Teachers were then asked to share out their thinking related to each group's title and decisions were made to combine groups or leave them the same, depending on unanimous agreement with all participants. Teachers struggled to limit their interaction with one another which further informs the researchers of the social nature of collaborative work and the need for flexibility within certain aspects of qualitative research. Teachers were given the opportunity to discuss titles and the decision to remove or change titles was only allowed when there was 100% agreement amongst the participants.

Once teachers finalized titles or names for each group (written on a different color sticky note) all participants were given a chance to vote on their favorite titles using three Xs to mark their highest priority themes/titles. These titles are then ranked based on the highest number of votes and listed in order of importance to represent the group consensus.

Analysis

In order to identify key areas of focus and opportunities for future prototyping for the MMaRS project, researchers set about working to better understand teachers' values and decision making around the experience of student assessment. MMaRS researchers assembled a team of human-centered design researchers (one emic and one etic observer to the project) and took time to develop a codebook for analyzing the data gathered from the fall and spring TAP research activities.

> **Researcher 1 (BAE).** Emic participant observer; embedded human-centered design researcher and Spring TAP meeting facilitator with elementary and early childhood education teaching experience.

> **Researcher 2 (B.S., M.Ed.).** Etic participant observer; educational coach/consultant and human-centered design researcher with secondary education teaching experience.

Phase I: Preliminary Analysis Process

The initial stage of data analysis within the TAP project included a complete overview of all data collected throughout year one and the development of key definitions aligned within the context that we are seeking to understand (see Codebook Appendix A). Researchers sought to identify prominent patterns or trends within the data and to better understand the categories of data that were collected throughout the first year. During the initial review of data, researchers categorized the sources of data into two types of prompts as seen in Table 2.

Open-Ended Prompts	Closed or Focused Prompts
K.J. Technique	Fall Focus Group
Love Letter/Break Up Letter Activity	"Dots" Exercise

Researchers then proceeded to work on establishing norms around the specific area of assessment development, instructional decision-making, and affective descriptors that would be coded throughout analyses. Working to develop a taxonomy, which serves to aid in the process of "deconstructing a situation into component parts and analyzing its aspects" researchers sought to find commonalities among the components that teachers described in their responses to the prompts (Shedroff 2003, p. 156). The coding of affective descriptors surfaced throughout the data review as a key element of the research that would be analyzed further to understand the inherent value that teachers' assign to any given aspect of the assessment experience. This led to the development of insights that will serve to direct next steps and future prototyping/iterations within the design cycle.

Researchers selected data collected during the Spring TAP focus group for thematic analysis. The data selected was from the K.J. Technique (affinity diagram) because the prompt given to the TAP teachers was open-ended. Researchers then employed thematic analysis in conjunction

with the taxonomy that was developed, to identify distinct components of the assessment cycle that teachers identified most frequently when prompted with a general, open-ended question:

What makes data useful and useable?

Based upon this initial synthesis; four primary areas of focus were identified: assessment design, the implementation of an assessment, how the results of an assessment are conveyed, and the instructional actions that teachers take. The definitions of each of these 4 areas of are shown in table 2 below.

Table 2

Design, Implementation, Results, and Actions Definitions

Design*	The foundational elements of the actual assessment. Both the form (structural) and function (purpose) of the assessment, as well as content (vocabulary, syntax, question types, etc.) are all considered as part of assessment design for our purposes. The nature or intent of the assessment, what it is intended to do is also included. For example, whether the assessment does what it says it is going to do (e.g. a math test that does not explicitly depend on students' reading ability to assess mathematical understanding). The design also includes designated goals or standards that the assessment is aligned to (I.e. CCCS or TEKs) and the district-adopted standards for practice. Design also includes whether the test is adaptive (does it increase or decrease in complexity based on student responses) and whether or not it is scripted for the proctor.
Implementation*	The execution of the actual assessment. This process includes, but is not limited to: what time of year the assessment is being used, whether that schedule is mandated or self-directed by the individual teacher, logistics of the arrangement of time and space for administration with the school setting, proctors, etc. Actual components utilized to assess which impact both teacher and student experience. For example; is it hands-on or computer-based? The type of language that is used (which can also be considered a part of the design; hence the overlap with teacher-facing materials and student-facing materials.
Results*	Information that is generated as a result of assessments that are given in the classroom. Any resulting data or raw scores that are associated with standardized or classroom-based assessment tools. This includes both progress monitoring tools and benchmark assessments. Results also include the issue of designations or categorizations that result from assessment tools; e.g. Beginning Reader (as defined by industry standards, number of words per minute etc.) Results are often reported out in some form and we include these "score reports" as a part of our definition of results when working with TAP. Results can be delivered electronically or in hardcopy.
Actions	Steps that are taken by teachers based on assessment results or information provided by any assessments administered in the classroom setting. Actions include, but are not limited to: instructional decisions, content and lesson design, interventions, technology implementation, grouping decisions, and any other practical arrangements made for instructional purposes (e.g. walk to math or reading groups) Actions, for the purpose of our research, also include decisions made that impact the student learning and assessment experience. These actions may be determined externally (E.g.: district or state mandated) or internally (E.g.: grade-level team or classroom-based.)

Note. *Area explicitly categorized by TAP Spring Affinity Diagramming

These areas include both the priorities identified by teachers and the key areas that emerged through the researchers' synthesis of data specifically from the KJ Technique method implemented during the Spring TAP meeting.

Researchers then participated in a preliminary tabletop sort of data from the Love Letters and Break Up Letters to norm within the four identified thematic categories utilizing data that was specifically affective in nature from the Love Letters and Break Up Letters. (see Codebook Appendix A). The process of analysis utilizing these four primary areas of focus was successfully applied and researchers were able to norm the code analysis and proceed to the secondary phase of analysis with the remaining TAP data.

It should be noted that through this initial phase of coding analysis, researchers determined similar patterns of "double-coded" responses that might suggest areas of focus which impact two of the given categories. These overlapping pieces of data resurface during the secondary phase of coding and contribute to the defining of both individual codes and themes throughout analysis. It is concluded that this is an inherent result of the cyclical nature of assessment and instruction which can be accounted in any further data analysis.

Phase II: Continued Analytic Process

Once the areas of overlap were defined and clarified, researchers collaborated to develop a codebook of codes and definitions for the overarching themes for thematic analysis, as well as, the rating scale for affective descriptors that would serve as a secondary layer of analysis. These codes were then refined after the first round of analysis with the Fall TAP data, as a result of data collected that represented teachers' feelings and thoughts related to reading curriculum and assessment.

The affective layer of analysis came in the form of coding for affect that was defined as positive, negative, or mixed. The operationalized definitions for these three areas of affect are shown in Table 3.

Table 3

Positive, Negative, and Mixed Affect Definitions

Positive	Something the teacher found beneficial, having an effect of bringing ease, help, or improving one's experience.
Negative	Something seen as increasing levels of frustration, disappointment, confusion or shame. Detrimental to accomplishing one's goal.
Mixed	Containing aspects of both positive and negative experiences, generating feelings of tension or discomfort as well as hope or wish-fulness. Resulting in hesitation and uncertainty.

A further refinement was made to account for the double-coding of mathematics and reading curriculum and assessment materials/experiences that emerged when reviewing data from the Fall TAP meeting. A future comparative analysis of reading and mathematics

curriculum/assessment experiences would provide applicable information for assessment development and design, however, due to the scope of this project, that data was omitted from the final analysis to ensure that the focus of this study remained exclusively mathematics-based.

With the code defined and refined to account for multiple sources, the research team then utilized NVivo to code each data source. This process allowed for coders to code each data source independently and analyze the coder agreements using an overall Kappa measurement embedded in NVivo. Researchers coded all data sources using the same seven categories. The resulting data conveyed information related to teachers' priorities with decision-making and actions around the assessment process, as well as more intrinsic indications of value related to teachers' affective expressions of emotions related to their experiences.

Phase III: Collective Analysis Process

The analyses plan was implemented in the coding of all of the data collected from the Fall TAP meeting and the Spring TAP meeting. This analysis included components of affective and objective data being coded to reflect what we are calling "value" as is reflected in teachers' use of a tool or item for a certain purpose and the overall experience of that use. For our purposes, value is defined as "having positive affect for and/or appreciation of a process, tool, experience, or person." Two coders analyzed transcripts and written data sources from both Fall and Spring TAP meetings to identify excerpts that reflected teachers' ideas and opinions about assessment and data. A sample of coded excerpts is highlighted in Table 4.

Table 4

Sample Coding Matrix of Excerpts

	Design	Implementation	Results	Action
Positive	*"So, the assessment is very place-value driven and so that's kind of nice."*	*"I love that you show up three times a year, rain or shine"*	*"Dear Assessment Data... I can depend on you."*	*"It [assessment data] is what helps me to know what to focus on."*
Mixed	*"We will assess each unit, but not ...we're just not putting as much weight in."*	*"It gave a lot of good information but it was way too long."*	*"I am losing too much instructional time by giving the assessment to every child, but it's very informative."*	*"Data should guide me, but is so often doesn't."*
Negative	*"Beating down my students with your long windedness and strange vocabulary so it appears they do not know things they do know."*	*"So, you gave up instructional time... I'd want to replace it with something you're doing because it can be the fourth test..."*	*"I don't feel like it is an accurate picture."*	*"I can see the frustration with my students and yet sometimes I have to keep pushing."*

Table 5

Inter-rater Reliability Scores for Assessment Data Analysis

Selected Nodes (Name)	Kappa	Agreement	Disagreement
Affect – MIXED	0.77	94.75	5.25
Affect – NEGATIVE	0.85	93.37	6.63
Affect – POSITIVE	0.85	94.75	5.25
Assessment Design	0.61	91.65	8.35
Implementation	0.86	95.05	4.95
Instructional Actions	0.70	92.83	7.16
Results	0.83	93.57	6.43

Overall unweighted kappa: 0.81

Findings

Fall TAP Meeting: Focus Group

Teachers responses varied when prompted specifically regarding instructional practices, the role of data in the classroom, and their experience with curriculum and assessment. The transcript from this focus group was coded in NVivo using not only the four assessment implementation nodes- assessment design, implementation, results, and instructional actions- but also, when applicable, three affective nodes: positive affect, negative affect, and mixed affect.

Through NVivo, we were able to determine the total number of references coded at each of the four assessment implementation nodes. Additionally, we were able to run a matrix coding query to determine the number of references double coded for affect at each node. The results of these queries are shown in Table 6.

Table 6

Fall Focus Group- Assessment Implementation and Affect Data

	Assessment Design			*Implementation*			*Results*			*Instructional Actions*		
Total number references coded	49			40			42			61		
Percent of total references coded	**25%**			**21%**			**22%**			**32%**		
	+	*Mixed*	-	+	*Mixed*	-	+	*Mixed*	-	+	*Mixed*	-
Number of affective references coded	5	11	12	2	8	11	3	8	10	5	5	10
Percent of references affectively coded	**10%**	**22%**	**25%**	**5%**	**20%**	**28%**	**7%**	**19%**	**24%**	**8%**	**8%**	**16%**

These data demonstrate that while TAP teachers referenced instructional actions the most frequently, overall, they referenced each of the four nodes with a relatively equal distribution

throughout the focus group. At each node, we can both see the number and distribution of responses that were coded affectively and calculate the number of responses that were "neutral", or coded without affect. Comparing the total number of references coded at each node to the number of references coded affectively at the same node reveal that an average of 46.9% of teachers' responses were double coded with affect while an average of 53.1% were not coded as affective. While the average number of responses coded affectively per node was 46.9%, the percent of responses coded affectively at each node varied: 57% for the assessment design nodes, 53% for the implementation node, 50% for the results node, and 33% for the instructional actions node. Analyzing the distribution of affective references demonstrates that, for each of the four nodes, teachers' responses were negative in affect at least twice as frequently as they were positive. In fact, in the references made to implementation, teachers' responses were coded as negative in affect 5.6 times more frequently.

Additionally, we were able to run a matrix coding query to determine the number of responses that were double coded for the four assessment implementation nodes. The results of this query are shown in Table 7.

Table 7

Fall Focus Group- Data for Associations Between Assessment Implementation Nodes

	Total references	References single coded	Double coded Assessment Design	Double coded Implementation	Double coded Results	Double coded Instructional Actions
Assessment Design	**49**	**69%** *(34)*		**29%** *(14)*	**31%** *(15)*	**14%** *(7)*
Implementation	**40**	**65%** *(28)*	**35%** *(14)*		**23%** *(9)*	**10%** *(4)*
Results	**42**	**64%** *(27)*	**36%** *(15)*	**21%** *(9)*		**29%** *(12)*
Instructional Actions	**61**	**63%** *(38)*	**11%** *(7)*	**7%** *(4)*	**20%** *(12)*	

	Total references	References single coded	Double coded Assessment Design	Double coded Implementation	Double coded Results	Double coded Instructional Actions
Assessment Design	**49**	**69%** *(34)*		**29%** *(14)*	**31%** *(15)*	**14%** *(7)*
Implementation	**40**	**65%** *(28)*	**35%** *(14)*		**23%** *(9)*	**10%** *(4)*
Results	**42**	**64%** *(27)*	**36%** *(15)*	**21%** *(9)*		**29%** *(12)*
Instructional Actions	**61**	**63%** *(38)*	**11%** *(7)*	**7%** *(4)*	**20%** *(12)*	

These data demonstrate that though teachers' responses were single coded a majority of time- an average of 66%, for all four nodes- they often spoke about their experiences in such a way as to reference two or more nodes in a single response. Analyzing the data from this table reveal the ways in which teachers referenced- or connected- nodes in their responses. For instance, these data indicate that teachers referenced instructional actions more frequently throughout the focus group than they did any other single node. However, the fact that this node was single coded 63% of the time- less often than any of the other three nodes- demonstrates that teachers referenced instructional actions in conjunction with one of the other three nodes more frequently than they did with regard to any other node. While the instructional actions node was the node that was most frequently double coded with some other node- most notably the results node- there were other nodes with higher frequencies of association: the two nodes with the highest association are the results and assessment design nodes, which were double coded in teachers' responses 15 times.

During the fall focus group, teachers were asked about the assessments and materials that they use- or have used- in their classrooms. During this time, teachers spoke at great length about the *design* of the math assessments, the *implementation* of these assessments, the communication of students' assessment *results*, and the teachers' *instructional actions*. Throughout this focus group, teachers' responses made reference to instructional actions more frequently than any of the other three nodes. Additionally, teachers' references to instructional actions were double coded more often than any other three nodes. We believe that this greater affinity for association with other nodes indicates that the TAP teachers saw instructional actions as the most connected of the nodes. However, while teachers' references to instructional actions were the most prevalent and most often associated with other nodes, they were also the least frequently affective.

To better understand what these aggregated quantitative data might mean, we looked at it in conjunction with the individual qualitative items coded to each of the nodes. The quote below was coded to instructional actions and encapsulates the characteristics captured in the data and descriptions above: it contains a lot of specific information about the teacher's instructional actions, but these instructional actions also make reference to other nodes in a way that is void of affect.

> Most of my groups are more of a reteach kind of a group. We do our own grade level created a Friday quiz that spirals. Our district provides an end of unit small assessments, not long, five, six, eight questions just to see if they've mastered a skill. It's not meant to be the end all and be all, but it's one data point, but as a grade level, we do a Friday quiz that spirals back, over everything. And then we get together and say, 'What is your kid still struggling with?' And so we'll put some of those questions again on our Friday, little Friday quiz that they do, maybe 15 questions, 20 just touches back on those skills we've been working on in small groups.

Select representative quotations for each of the other node are shown in Table 8.

Analyzing qualitative data, such as the above quotations, coded to each of the nodes alongside the aggregated quantitative data revealed some interesting themes. One theme that we noticed across all nodes was an implicit and explicit reference to the teachers' *agency*. Here, we are defining agency as the capacity of an individual to act independently and/or to make their own choices. Further, we deduced that agency was often correlated with the teacher's affective response, particularly negative affect. We noticed that when a teacher's response implicitly or explicitly conveyed that their agency was diminished in some way, their response was often negative in affect. Contrastingly, when a teacher's response implicitly or explicitly conveyed that their agency was expanded in some way, their response was less likely to be negative in affect. In these instances, statements may have been conveyed in such a way that they were coded with positive affect, mixed affect, or were not coded for affect at all ("neutral").

Table 8

Fall Focus Group- Matrix of Sample Quotations

Assessment Design	**Negative Affect:** *"Our UbDs are like that too, time and money or never assessed. And when they get to second they do not know them. They do not know them because they're not being assessed."* **Positive Affect:** *"And we have, access to ST math, which I really like because I could put kids into kindergarten or second grade or third grade or first whatever they need. And so that, that helps with the little ones that need the repetitive practice."*
Implementation	**Negative Affect:** *"The math Istation. My thing is, it starts them all at the same thing where they're all counting one, two, three, four. Instead of putting them where they're at, like the reading does in Istation. You know, it doesn't start them all at the same beginning lesson."* **Mixed Affect:** *"Because there's more than an hour per kid probably like there's like 15 different section lists long, because I'm giving it to second graders who are proficient on more of it. So like kindergarten says it doesn't take them very long because they top out ...when I get them in second grade, I should only be picking up on the things that I know that they don't have, you know, or maybe going back one step. So it shouldn't take as long, but I foresee that they will stop doing it because it's just, it takes too long. It's, I'm losing too much instructional time by giving the assessment to every child, but it's, very informative. It's very informative."*
Results	**Negative Affect:** *"Those are the things that take time though, because this other assessment I'm talking about every answer I have to say, "How did you get that? How did you get that? How did you get that?" And then I have to write down what they say and that's really time consuming, you know, or I have to circle and I have to know what all these letters mean because... [there are] so many ways to get an answer. So that's kind of tricky."* **Mixed Affect:** *"I can't remember what that one was called. But anyway, it's more like ... They would always say it's like a net to see, you know, where they're at, but it wasn't helpful in the sense of like how to group kids or what they're missing or anything like that. So, their primary numeracy is helpful because it does show you where those gaps are. And so we can use that information, like I was just explaining. So now we've, changed those intervention groups to a math intervention group and we're using like the counting forward, counting backwards and counting by 10s on and off decade. We're using that information to get our kids where they need to be from that assessment."*

We believe that this theme relates to one of the main trends in the quantitative data presented in Tables 4 and 5: teachers' references to instructional actions were the most prevalent and most often associated with other nodes, but were also the least frequently affective. We believe that the significant difference in the frequency of affective references between instructional actions (33% affectively coded) and the other three nodes (average of 54% affectively coded) could relate to agency. Teachers references to assessment design, implementation, and results more frequently suggested that their agency was diminished in some way; teachers seem to feel as though they have less control over these nodes or that these nodes result in them having diminished agency in some way.

Spring TAP Meeting: Dot Exercise

The dot exercise is a "heat mapping" method for collecting individual, quantifiable data related to specific statements. The specific statements that we sought to collect data regarding were key themes and insights that were extracted from the fall focus group data. The insights centralized around how data is used to make decisions for grouping and next steps in instruction and planning. These statements were presented to participants on three prepared posters, each with nine statements. The centralized question presented to participants on each poster varied, and the statements that participants were to rank ranged in content from grouping decisions to instructional data use. Upon arrival, participants received three sets of three green and three red "dots", one set (6 dots: 3 green, 3 red) for each poster. Teachers were then prompted to apply the dots according to their level of priority for statements on each of the posters, using six dots per poster. Green dots indicated high priority and red dots indicated lowest priority. The combination of dots per poster reflected participants agreement or disagreement with the prioritization of each given statement. No dot indicated no prioritization or consideration.

The data from this exercise existed in the form of red and green dots aligned to poster statements and was manually coded for the four assessment implementation nodes- assessment design, implementation, results, and instructional actions- by tabulating the quantity of dots aligned to each statement. The results of this coding exercise are shown below in Tables 9-12.

Table 9

Dot Exercise: Prompting Statement Color Coding Key

Assessment Design
Assessment Implementation
Assessment Results
Instructional Actions

The key above shows the color coding that will be used in Tables 10-12. This color coding will be applied in the columns labeled 'prompting statement' in the tables below. The color coding applied to the statement reflects the node to which the statement aligns.

Table 10

Dot Exercise: "How do you use data to plan the "next steps" in your instruction?"

Highest Priority Statements		
Prompting Statement	Green Dots Placed	Percent of Green Dots
Determine which content should be weighted more heavily in instruction	7/18	39%
Know how to increase the sophistication of content	4/18	22%
Understand what content needs to be spiraled	4/18	22%
Lowest Priority Statements		
Prompting Statement	Red Dots Placed	Percent of Red Dots
Know which instructional strategies are most effective	12/18	67%
Identify which content should be the basis of assessments	4/18	23%

Table 11

Dot Exercise: "Based on what characteristics do you want to group students in your math instruction?"

Highest Priority Statements		
Prompting Statement	Green Dots Placed	Percent of Green Dots
Students' current level of skills and knowledge	6/18	33%
Gaps in students' understanding based on errors or misconceptions	5/18	28%
Lowest Priority Statements		
Prompting Statement	Red Dots Placed	Percent of Red Dots
Students affective characteristics	8/17	47%

Table 12

Dot Exercise: "How will you use grouping information?"

Highest Priority Statements		
Prompting Statement	Green Dots Placed	Percent of Green Dots
Create groups to use in the workshop model of instruction	6/18	33%
Identify gaps in students' understanding	5/18	28%
Lowest Priority Statements		
Prompting Statement	Red Dots Placed	Percent of Red Dots
Create different problem sets for students to complete in class or as homework	7/18	39%
Organize number talks	6/18	33%

Teachers responses to the Dot Exercise included a consistent pattern of prioritization of grouping decisions and targeted skills for instructional actions. These data speak to the research question "what are key areas of application to prioritize when designing tools to make data useable for practitioners?"

Spring TAP Meeting: Love Letters and Break Up Letters

Teachers were given a "Dear Assessment Data," letter prompt and asked to construct both a "love letter" and a "break up letter". Each teacher produced two unique documents: one "love letter" document and one "break up letter" document. All of the documents from this activity were coded in NVivo using not only the four assessment implementation nodes- assessment design, implementation, results, and instructional actions- but also the three affective nodes: positive affect, negative affect, and mixed affect.

Through NVivo, we were able to determine the total number of references coded at each of the four assessment implementation nodes. Additionally, we were able to run a matrix coding query to determine the number of references double coded for affect at each node. The results of these queries for the aggregated data from both the love letters and break up letters are shown in Table 13.

Table 13

Aggregated Assessment Implementation and Affect Data from Love and Break Up Letters

	Assessment Design			*Implementation*			*Results*			*Instructional Actions*		
Total number references coded	19			43			47			22		
Percent of total references coded	**15%**			**33%**			**36%**			**17%**		
	+	*Mixed*	-	+	*Mixed*	-	+	*Mixed*	-	+	*Mixed*	-
Number of affective references coded	4	2	13	8	1	34	21	8	18	9	3	10
Percent of references affectively coded	**21%**	**11%**	**68%**	**19%**	**2%**	**79%**	**45%**	**17%**	**38%**	**41%**	**14%**	**45%**

These data depict the total number of references that TAP teachers made in both love letters and break up letters. These data demonstrate that while TAP teachers made reference to all four assessment implementation nodes, they referenced the implementation and results nodes more than twice as often as they did the assessment design or instructional actions nodes. At each node, we can both see the number and distribution of responses that were coded affectively and calculate the number of responses that were "neutral", or coded without affect. Comparing the total number of references coded at each node to the number of references coded affectively at the same node reveal that 100% of teachers' responses were double coded with affect, during this activity. Analyzing the distribution of affective references for each of the four nodes shows that the net affect with regard to each node is quite different. First, comparing each node's percent positive to percent negative affect reveals that results was the only node with a greater weight of positive than negative affective references: teachers' references to results were coded as positive in affect 1.2 times more frequently than they were coded as negative in affect. Contrastingly, the

remaining three nodes had a greater weight of negative than positive affective references: teachers' references to instructional actions were coded as negative in affect 1.1 times more frequently than they were coded as positive in affect; teachers' references to assessment design were coded as negative in affect 3.3 times more frequently than they were coded as positive in affect; and teachers' references to implementation were coded as negative in affect 4.25 times more frequently than they were coded as positive in affect.

While the aggregated data give us insight into the distribution of references in teachers responses overall, the distribution of data is vastly different when analyzing the love and break up letters separately. Through NVivo, we were able to run separate matrix coding queries for each source- all of teachers' love letters in aggregate and all of teachers' break up letters in aggregate- in order to determine the total number of references coded at each of the four assessment implementation nodes as well as the three affective nodes. The results of the query for the love letters is shown in Table 14 while the results of the query for the break-up letters is shown in Table 16.

Table 14

Love Letter Assessment Implementation and Affect Data

	Assessment Design			*Implementation*			*Results*			*Instructional Actions*		
Total number references coded	6			9			27			11		
Percent of total references coded	**11%**			**17%**			**51%**			**21%**		
	+	*Mixed*	-	+	*Mixed*	-	+	*Mixed*	-	+	*Mixed*	-
Number of affective references coded	4	2	0	8	1	0	21	6	0	9	2	0
Percent of references affectively coded	**66%**	**33%**	**0%**	**89%**	**11%**	**0%**	**78%**	**22%**	**0%**	**82%**	**18%**	**0%**

The love letter data shown in the table above reveal that teachers' responses referenced results significantly more often than they did the other three nodes: the results node was referenced more times than were the other three nodes combined. As was noted in the discussion of the aggregated data from table 11, 100% of responses were double coded with affect. And while a majority of the references for each node were coded as positive in affect- and none of the references were coded as negative in affect- each of four nodes contained references that were coded as mixed in affect. Representative quotations for each node are shown in Table 15.
The sample quotations in Table 15 are snapshot of the responses provided by teachers in their love letters. Analyzing all of the content coded to each node provides insight into what teachers most value about assessment data from their own experience in their current practice. Looking at the data coded with positive affect across all four of the assessment implementation nodes, we deduced some overarching themes: teachers' responses demonstrated that they valued consistency and predictability as well as the data's ability to empower strategic action. As a matter of fact, in their love letters, all teachers made reference in one way or another to the importance of data conferring the ability to empower them to make strategic decisions or actions in order to better serve students. These themes are further supported when looking at those quotations coded for mixed affect. Here, we noticed a consistent tone of skepticism. Further, teachers' responses implied a lack of trust, particularly with regard to the assessment data's

ability to provide information in such a way as to empower their ability to make strategic instructional decisions.

Table 15

Love Letter- Matrix of Sample Quotations

	Positive Affect	Mixed Affect
Assessment Design	*I can depend on you.*	*You say you've changed. Have you?*
Implementation	*I love that you show up 3 times each year, come rain or shine.*	*Dare I hope that you won't turn against me. Let's try one more time.*
Results	*You are so great when you give me detailed information that shows how my students are thinking.* *I am then able to compare your previous visits with your real-time visit so I can see where my students stand.* *Also, it is lovely when I see patterns with groups of students and therefore can give a teaching point that is effective to a small group of students.* *I really needed you to remind me that Doris doesn't always remember how to subtract and she sometimes needs reminders to help get her on track.*	*Will kids be served by the information you hold?* *I love the detailed information some of you (Primary Numeracy Assessment) provide.*
Instructional Actions	*It makes it easy for me to group my kids and also to know what skills I need to reteach.*	*Here you stand... promising to show me the path forward. How I wish I believed this prosaic story.*

There were both similarities and differences in the data gathered from teachers' break up letters. The data from the break up letters are shown in Table 16.

Table 16

Break Up Letter Assessment Implementation and Affect Data

	Assessment Design			*Implementation*			*Results*			*Instructional Actions*		
Total number references coded	13			34			20			11		
Percent of total references coded	**17%**			**44%**			**26%**			**14%**		
	+	*Mixed*	-	+	*Mixed*	-	+	*Mixed*	-	+	*Mixed*	-
Number of affective references coded	0	0	13	0	0	34	0	2	18	0	1	10
Percent of references affectively coded	**0%**	**0%**	**100%**	**0%**	**0%**	**100%**	**0%**	**10%**	**90%**	**0%**	**9%**	**91%**

The break up letter data shown in the table above reveal that teachers' responses referenced implementation significantly more often than they did the other three nodes. As was noted in the discussion of the aggregated data from table 11, 100% of responses were double coded with affect. And while an overwhelming majority of the references for each node were coded as negative in affect, two of the nodes, results and instructional actions, contained references that were coded as mixed in affect. Representative quotations for each node are shown in Table 17.

Table 17

Break Up Letter- Matrix of Sample Quotations

	Sample Quotations
Assessment Design	*Beating down my students with you long windedness and strange vocabulary so it appears they do not know things they do know.* *Just leave me alone and let me ask what I want and need to know.* *You say one thing and do another.*
Implementation	*Why did you pick today to come by? Didn't you know that Mason didn't take his meds this morning?...What about Felicity? Did you not hear her fighting with Gabby at recess and then cry in the corner of the room? You are no good to me today- GO AWAY!* *Just when I get done with you, you show up again!* *I always want to break up with you when you drag on and on. I can see the frustration with my students and yet sometimes I have to keep pushing. Often, I feel like all of my time is devoted to assessing, and I never can get to the point where I can TEACH. I'm just so exhausted from testing 22 kids and it taking so long.*
Results	*It is even more frustrating when the data obtained seemingly has nothing to do with what I teach or need to teach.* *Sick of you wasting my time and energy only to give me vague, unspecific information (Imagine Math)* *The emotional damage you bring to kids is unforgiveable. We are more than a test score!* *So many times your information is not accurate. My kids know so much more than your info shows. I'm done!*
Instructional Actions	*Data should guide me, but so often it doesn't.* *I'm tired of pouring all of my time and energy into you so I have nothing left for what matters- actually teaching my students.*

The sample quotations in Table 17 are snapshot of the responses provided by teachers in their break up letters. Analyzing all of the content coded to each node provides insight into what teachers most value about assessment data from their own experience in their current practice.

Looking at the data coded with negative affect across all four of the assessment implementation nodes, we deduced some overarching themes: teachers' responses demonstrated that they valued alignment and validity, maximizing and honoring teachers' time, and the ability to empower strategic action. In their break up letters, all teachers made reference in one way or another to the importance of data accurately reflecting and reporting the true nature of their students' achievement. Teacher discussed how assessment data may not provide an accurate picture of achievement by addressing various root causes for assessments failing to provide data that is valid, useful, and usable: some teachers commented on the design of the assessment, particularly in the way of alignment); others remarked on assessment implementation, particularly with regard to failing to honor teachers' and students time and demonstrating a lack of reverence for the true nature of teacher and student experience when being assessed; and many discussed the lack of usability of assessment reporting or results. Though teachers discussed the importance of data accurately reflecting and reporting the true nature of their students' achievement in various ways, the implication in teacher's responses seemed to be that they valued this for two reasons. First, as was mentioned in the discussion of quotations from love letters in table 15, having valid and accurate data confers the ability to empower teachers to make strategic decisions or actions in order to better serve students. Second, teachers recognize the importance of considering humans to which the data relate and individualizing for unique contexts.

As was mentioned in the definition given for this activity in the method section, the intention of the Love Letter/Break-up Letter method helps designers "understand the less tangible aspects of the things they create; specifically, the social, human values and meanings conveyed through the things and experiences we design, as well as their understandability and usability." (Shedroff 2003, p. 159) Through this method we provided an open ended prompt and freedom of expression in order to understand what teachers value about assessment data by probing for the concepts that were most salient to teachers, both in the things they appreciate and the things that they do not, in their current practice and use of assessment data. We believed that the nature of the content in their letters would demonstrate what they most valued about assessments, as they currently experience them.

Teachers' responses in writing to Assessment Data during this exercise reflected a complex relationship between data and assessment. Teachers' expressed affect and comments that reflected a value for effective assessment data. They also expressed discontent and frustration in the form of negative affect and commentary regarding assessment implementation. Analyzing qualitative data, such as the quotations in Tables 15 and 17, coded to each of the nodes alongside the aggregated quantitative data revealed some interesting trends that provide insight into what TAP teachers value about assessment data. Two trends that we noticed across all nodes were implicit and explicit references to the ability of assessments to be adaptable and individualized for teacher and student context as well as the ability of assessments to empower teachers to make strategic decisions or actions in order to maximize time and best serve students. In expressing both of these concepts, teachers conveyed the importance of intentionally creating alignment (from the content for which teachers are accountable to the design of an assessment to the means by which the assessment is implemented to the way in which students' achievement are conveyed in results) such that data is valid, useful, and usable.

Spring TAP Meeting: KJ Technique

When prompted with the open-ended prompt "What makes data useful and useable to you?" A facilitator guided the group through independently brainstorming ideas that were most salient to them. Participants were told that they could write down as many ideas as they'd like to independently, but asked individuals to place only one concept on each post-it. Teachers' responses varied with a wide range of consideration of every aspect of the process of assessment and data collection. Teachers were concerned with everything from how the test is written (vocabulary and content) to how assessment data is generated and communicated to students, families, and administrators (score reports). The facilitator then guided the participants through the process of grouping and regrouping ideas silently, then through the step of identifying themes and creating names or titles for each group, first individually, then collectively. Once teachers finalized titles or names for each group (written on a different color sticky note) all participants were given a chance to vote on their favorite themes/titles using three Xs to mark their highest priority, two Xs to mark their second highest priority, and one X to mark their third highest priority. These titles were then ranked based on the highest number of votes and listed in order of importance to represent the group consensus.

The data from this exercise existed in the form of post-it notes and was manually coded for the four assessment implementation nodes- assessment design, implementation, results, and instructional actions- by sorting the post-it notes that teachers produced. A table showing a written record of the manual post-it note coding from the KJ technique is provided in Appendix B. Through manual coding, we were able to determine the total number of references coded at each of the four assessment implementation nodes; the results of this coding exercise are shown in Table 18.

Table 18

KJ Technique Assessment Implementation Data

	Assessment Design	*Implementation*	*Results*	*Instructional Actions*
Total number references coded	38	15	17	22
Percent of total references coded	**41%**	**16%**	**18%**	**24%**

As was mentioned in the definition given for this activity in the methods section, the intention of the KJ technique is to encourage "consensus building that is silent, makes effective use of time, and negates "group pressure" by providing personal representation of perspective with minimal judgement passed within a group setting." (Kawakita 1982) Through this method, we provided an open-ended prompt "What makes data useful and useable to you?" in order to understand what teachers value about data, mapping back to the research questions. We believed that the nature of the content that they would produce on Post-Its would not only demonstrate what they most valued about data, but also and reveal how they conceptualized data implementation in an ideal world.

The data from Table 16 reveal that while TAP teachers made reference to all four assessment implementation nodes, teachers' responses referenced assessment design significantly more often than they did the other three nodes. The main areas of interest for the TAP (highest priority in voting) and the areas that were coded with the most numbers of entries overlapped consistently. Teachers developed their own agreed-upon categories to delineate how data can be useful and useable. These categories are Design, Administration, and Results. Within these categories, the TAP prioritized three main areas of focus: Alignment, Administration, and Classroom Instruction.

What is most interesting about this data is the picture that it paints when triangulated with the data collected though other methods from TAP teachers. During the KJ Technique exercise, researchers implicitly asked teachers to consider an ideal world by focusing teachers on expressing positive and ideal attributes with the "what makes (or would make) data useful and usable". Asking teachers to think in this asset-based way provides a solutions orientation and reveals a prioritization of the design of the assessment itself, which is distinct in comparison to the prioritizations depicted in data from previous collection activities. From other exercises, teachers' responses reflecting the current state of assessment implementation demonstrated that, overall, teachers more frequently referenced areas with negative affect than they did with positive affect, especially in relation to assessment implementation. Additionally, we saw that teachers value agency and an assessments ability to empower them to make strategic decisions or actions to best individualize for students to meet their needs. Finally, in other sources probing teachers current experience, we noted that teachers deeply valued consistency, transparency, and alignment with regard to all assessment implementation nodes. Those trends in conjunction with the data from the KJ technique demonstrate the extreme importance of thoughtful, intentional assessment design with an eye toward flexible implementation and individualization as well as the communication of results that are clear, actionable, and aligned to meaningful content.

Insights

Findings from the first year of TAP research provide valuable insights about our primary research question: How might we develop assessment resources that are valuable to teachers in the classroom?

Actionable Data are Ideal

Teachers value data and, in an ideal world, want it to be an engine for their instructional actions.

The TAP teachers expressed a tremendous value for data that was directly linked to their students' learning and areas of growth. Teachers are interested in providing instruction that targets explicit skills and gaps in understanding and the more targeted the assessment results are, the more refined learning targets can be to meet student need. Connecting assessment data directly to instructional practice is the most important theme that we identified across our triangulated data sources. Teachers made reference to their grouping decisions and instructional actions in the fall and emphasized the value and importance of data that is directly correlated

with their student learning outcomes. Having data that is interpretable and actionable empowers teachers to make instructional decisions that they feel confident will meet their students' needs effectively.

Affect is Correlated with Agency

Teachers' agency within a district greatly impacts their ability to adapt areas of assessment implementation and instructional action to meet their students' needs.

Based on our findings, the TAP teachers' experiences often reflected feelings of powerlessness and frustration related to their ability to do what was "best" for their students. In this context, there were multiple instances of different participants reflecting on the conflict they experience related to both the design and the implementation of assessments. These feelings of powerlessness also impacted their ability to make instructional decisions in the classroom setting and we believe that "best practice" instructional actions were impeded based on an externalized locus of control for some teachers.

Measuring student growth and achievement is an indispensable practice in the cycle of goal-setting and instructional practice in the classroom. The value of the information that can be collected with consistent and effective data collection instruments can positively impact the learning experience of all learners in a classroom, as well as lead to increased achievement over time. (Lambert, Algozzine, and McGee, 2014) A variation exists across districts, as well as within districts at times, of the practice of mandated, district-adopted benchmark assessments and the impact of this variation is reflected in teachers' affect related to assessment. However, the external mandates placed on practitioners have been evident throughout the data that we collected here. Increasingly more demands are placed on teachers' schedules. As a result of the tension in teachers' decision-making related to time management and best-practices, teachers who experience varying degrees of agency within the classroom have a wide range of feelings in relation to the use of assessment and data. This insight points to the need for further research in relation to the effects of agency and empowerment of teachers' decision-making and student growth.

Agency in Implementation Matters

Teachers feel restricted with current state of implementation.

The TAP data revealed another insight related to the actual implementation of assessments and teachers' experience with constraints within the process. When describing their ideal relationship with assessment data, teachers often pointed to the "just what I needed" function of data and assessment. TAP teachers' value for data that is valid and confirming is evident in our research. Often teachers described assessments that gave them the information they needed to know, as well as understood their students' needs on a given day to not be assessed due to a current family crisis or health concern. Teachers expressed an appreciation for flexibility in implementation and a sense of empowerment when districts accounted for teachers' formative assessment of learners in conjunction with the mandated assessments.

Trusting that teachers know their students' needs in a way that cannot be measured by an inanimate tool is a potential component of releasing the restriction that teachers feel in the implementation of assessments. The degree to which teachers are empowered to have agency within assessment cycles provides a further opportunity for investigation that is outside the scope of this research.

Assessment Results Impact Actions

Assessment results must not only be aligned to what teachers and students are held accountable for, but must also be conveyed in a way that is clear and actionable.

TAP teachers communicated a strong desire for data/reports that were easily interpretable and actionable in the classroom setting. Multiple references were made to data that was confusing and reports that were not easy to communicate the results from to students or families. Teachers desire assessment reports or results that translate into action effectively and conveniently in the classroom setting. Making decisions based on data and score reports is a regular practice in all of the classrooms of the TAP participants.

The references that teachers made to assessment results and the need for reporting tools that directly translate into the classroom (as a goal setting tool), the report card or standards-based reporting system (for communicating with families), and the administrative setting for further intervention and progress monitoring purposes, were plentiful. The specific needs of each of the five districts varied, based on the available resources, but across the TAP, teachers agreed that effectively communicated results are of tremendous value to classroom teachers.

Limitations

Sample Demographic/Size

The TAP committee included eight teachers from public school districts from areas surrounding the Dallas Metroplex. One public charter school teacher also participated in the TAP. Teachers who were recommended represented an average of 19 years of teaching experience. This presents both as a limitation as well as an asset when considering the implications of years of experience accrued in practice, as well as the degree of skill that has been developed by these teachers in the application of assessment data. The perspectives represented by teachers can be considered reflective of teachers who continue to strive for best-practices, yet consistently find vacuums within the present mathematical assessment landscape for primary grades.

The demographics of the student populations within the districts and the size of the districts that are represented by the MMaRS TAP participants reflects a collective sample of the urban and rural districts in and around the Dallas-Metroplex. The teachers themselves identify as being of Caucasian descent. The inter-suburban communities that the majority of these districts serve include a relatively less diverse population of students to that which is reflected in the greater metroplex. These demographic data of the districts are depicted in Table 19.

Table 19

Demographic Data of Districts

School District	Charter School District	Major Suburban	Other Central City	Other Central City Suburban$_1$	Other Central City Suburban$_2$
Total Enrollment*	19,000	13,957	29,952	3,039	2,684
Demographic Make-up	6% White 65% Hispanic 16% African American 11% Asian 2% Other	54.93%White 24.6%Hispanic 5.38%African American 9.23% Asian .43% Native American, Pacific Islander 5.43% Two or More	79% White 19.5%Hispanic 9.5% African American 3.4% Asian 4.8% Other	98.2% White 7.9% Hispanic .1% African American .3% Asian .8% Other	62% White 22% Hispanic 8% African-American 3% Asian 5% Other

*TXschools.gov (Texas Education Agency)

Time of Year of Data Collection

Focus on data increases as the academic year comes to an end. Teachers met with us for our spring focus group either during or at the end of their End of Year (EOY) testing cycles. Participants' affect and focus may reflect patterns relative to this time of year and assessment cycle.

Conclusion

The TAP component of our research afforded a valuable and important opportunity to apply Human-Centered Design methodologies to the research and development process of the MMaRS project. The insights gained through this process provide actionable and relevant direction for the next steps in this work. The teachers who participated in this research have communicated consistently and clearly regarding the need to consider the four major areas of the assessment experience (assessment design, implementation, results, instructional actions). Their participation has also provided us with feedback related to the greater consideration of agency and empowerment in the cycle of assessment and instruction. Teachers have conveyed their value for data, as well as opened a window into the practical application and function that data serves in K-2 classrooms. As indicated in our findings, teachers benefit from assessment tools that produce results which facilitate the decisions that need to be made for instructional actions in the classroom. There is, however, a need to better understand the relationship between how data is communicated and the variation in teachers' years of service and experience level. Assessments that provide data that are useful and useable for teachers at any stage of their career have the

potential to positively impact learning and assessment for all learners. Making data actionable for teachers begins with the development of assessment items and tools that account for the end-user at every step of the process. In the process of Human-Centered design, this accountability is embedded in the continuous feedback loop of iteration and prototyping. The TAP component of our work provides an opportunity for both future prototyping, and the eventual co-creation or participatory design of assessments. By utilizing design research methodologies and Human-Centered Design principles throughout the assessment development process we can elicit deeper levels of understanding into the authentic experience of educators with assessment and data in the K-2 classroom and produce increasingly useable tools

Although, traditionally K-2 classrooms are not considered "high stakes" testing environments- the degree of feeling and level of commitment to student achievement that was emphasized throughout the first year TAP data tells a different story. The "high stakes" that we see at the K-2 level include the instructional decisions that are being made as a result of the data that is collected through assessments that are implemented. There are also the affective experiences of the teachers being asked to (or electing) to assess learners with rigor and professionalism in K-2. TAP teachers provided us with insights into the process of assessment and data use in the classroom that illustrate the value systems and decision-making processes that data and assessment impact every day

Our next step recommendations include continuing to include teachers in the process of assessment development, including the value of learning progressions as a part of the assessment. Additionally, we recommend considering the value of including other end users including district coordinators, administrators, and data or instructional coaches in the development process. Moving into the next phase of prototyping around effective assessment tools could include elements of "Show me your most effective assessment tool? What are teachers really using in their classroom?" We also recommend exploring with practitioners the specific areas of the assessment process that were defined and highlighted in the TAP data this year and digging deeper into the areas of results and instructional actions.

Some areas of opportunity for more effective design include:

1. Keep the end-user in mind throughout the assessment design process...
    - What interpretations can I make from these assessment results?
    - What does the data tell me about this student's ability to perform a given task? Does it apply a particular concept?
    - What action does this data direct me to make in order to best meet my student's needs?

2. Be mindful of how results are conveyed to students, families, and teachers. Aligning results with actionable steps aids implementation.

- Conduct an inventory of score reports.
- Have TAP teachers bring in their favorites/most useful score reports, and the ones they'd like to burn to a crisp.
- Ask for feedback on assessment and report designs, early and often.
- Invite other end- users to provide feedback or design their own assessment report.
- Give items writers an opportunity to understand the experience of assessment implementation. Conduct mock "assessment experience" sessions- best case and worse- case scenarios.

The wide variance in data management systems, coupled with the complicated nature of the assessment/instructional cycle provide ample opportunities for improvement in design and implementation that can benefit from firsthand experiences and case studies as to how data management systems are implemented, maintained, and utilized by educators in the classroom. If we as researchers can ensure that the design phase of assessment development accounts for the value that teachers have for data and their need to take action with that data to positively impact student learning in the classroom, then the effectiveness of our assessment tools will have the staying power to ensure learners' needs are continuously addressed and teachers continue to value the useable data that: "*helps direct their steps and light their way*" - participant's Dear Assessment Data Love Letter. Identifying these themes and corresponding insights from the Understand and Define phases of the HCD process will inform our next steps and recommendations as we move into the ideation and prototyping phase of the MMaRS research with the TAP.

References

Bowie, A., & Cassim, F. (2016). Linking classroom and community: A theoretical alignment of service learning and a human-centered design methodology in contemporary communication design education. *Education as Change, 20*(1), 126-148. https://doi.org/10.17159/19479417/2016/556

DeCuir-Gunby, J. T., Marshall, P. L., & McCulloch, A. W. (2011). Developing and Using a Codebook for the Analysis of Interview Data: An Example from a Professional Development Research Project. *Field Methods, 23*(2), 136–155. https://doi.org/10.1177/1525822X10388468

Giacomin, J. (2014) What is Human Centered Design? *The Design Journal, 17*(4), 606-623. https://doi.org/10.2752/175630614X14056185480186

Glaser, R., Chudowski, N., & Pellegrino, J. W. (2001). In Glaser R., Chudowsky N. and Pellegrino J. W. (Eds.), *Knowing what students know: The science and design of educational assessment.* Washington, D.C.: National Academies Press. Retrieved from http://ebookcentral.proquest.com/lib/southernmethodist/detail.action?docID=3375251

Hanington, B. (2003). Methods in the making: A perspective on the state of human research in design. *Design Issues, 19*(4):9–18.

IDEO.org. (2015). *The field guide to human-centered design*. San Francisco, CA: IDEO.org. Retrieved from https://www.designkit.org/resources/1

Lambert, R. & Algozzine, B. & Mcgee, J. (2014). Effects of progress monitoring on math performance of at-risk students. *British Journal of Education, Society and Behavioural Science, 4*(4), 527-544. 10.9734/BJESBS/2014/7259.

Kawakita, J. (1982) *The Original KJ Method*. Tokyo: Kawakita Research Institute.

Martin B. & Hanington, B. (2012) *Universal Methods of Design*. Beverly, MA: Rockport

Merriam-Webster. (n.d.). Focus group. In Merriam-Webster.com dictionary. Retrieved April 24, 2020, from https://www.merriam-webster.com/dictionary/focus%20group

Shedroff, N. (2003) Research Methods for Designing Effective Experiences. In B. Laurel (Ed.) *Design research: Methods and perspectives (*pp. 156, 158-159). Boston, MA: Massachusetts Institute of Technology.

Southern Methodist University. (2020). DSIN 7303 – Human-Centered Design. *2019-2020 Graduate Catalog.* Retrieved from https://catalog.smu.edu/search_advanced.php?cur_cat_oid=41&search_database=Search&search_db=Search&cpage=1&ecpage=1&ppage=1&spage=1&tpage=1&location=33&filter%5Bkeyword%5D=human-centered+design&filter%5Bexact_match%5D=1

Weingarten, R. (2012). The role of teachers in school improvement: Lessons from the field. *Harvard Law & Policy Review, 6*(1), 9-38. Retrieved from https://heinonline.org/HOL/PrintRequest?handle=hein.journals/harlpolrv6&collection=0&div=4&id=9&print=section&sction=4

Appendix A –Teacher Advisory Panel Codebook

Purpose of TAP
How might we develop an [*assessment and corresponding learning progression*] that is valuable to teachers in the classroom?

Research Questions

What are key areas of application to prioritize when designing tools to make data useable for practitioners?

How do teachers use data in their communication about assessments?

What makes data useful in instructional design?

What areas of instructional design do teachers use data for?

What are the struggles that classroom teachers encounter in using data?

What are the wins that classroom teachers encounter in using data?

What factors influence teachers' use of data in the classroom?

How do teachers want to use data?

What role does data play in teachers' day-to-day decision-making?

What makes data useful to classroom teachers?

What makes data useable to classroom teachers?

Codebook parameters:

Define terms

Term	Definition
Useful	Assists in accomplishing a task
Useable	Assists in accomplishing a task as simply as possible, enjoyable and effective

Data	Raw scores resulting from assessment of students
Assessment	Formal and informal tools used to measure students' academic abilities and growth in the classroom. Can be summative or formative, can be classroom-based, district-adopted, or nationally normed. Some are mandated by district standards, some are self-selected by teachers.
Teachers	Elementary general education classroom teachers
TAP (Teacher Advisory Panel)	Eight teachers recommended to participate in focus groups with RME to discuss instructional practices, assessment, and data.
Value	Positive affect for and/or appreciation of a process, tool, experience, or person.

Focus Group Codebook

Overarching Thematic Codes (Targeted Areas of Practice)

Design	Implementation	Results	Actions
The foundational elements of the actual assessment. Both the form (structural) and function (purpose) of the assessment, as well as content (vocabulary, syntax, question types, etc.) are all considered as part of assessment design for our purposes. The nature or intent of the assessment, what it is intended to do is also included. For example, whether the assessment does what it says it is going to do (e.g. a math test that does not explicitly depend on students' reading ability to assess	The execution of the actual assessment. This process includes, but is not limited to: what time of year the assessment is being used, whether that schedule is mandated or self-directed by the individual teacher, logistics of the arrangement of time and space for administration with the school setting, proctors, etc. Actual components utilized to assess which impact both teacher and student experience. For example; is it hands-on or computer-based? The type of	Information that is generated as a result of assessments that are given in the classroom. Any resulting data or raw scores that are associated with standardized or classroom-based assessment tools. This includes both progress monitoring tools and benchmark assessments. Results also include the issue of designations or categorizations that result from assessment tools; e.g. Beginning Reader (as defined by industry standards, number of	Steps that are taken by teachers based on assessment results or information provided by any assessments administered in the classroom setting. Actions include, but are not limited to: instructional decisions, content and lesson design, interventions, technology implementation, grouping decisions, and any other practical arrangements made for instructional purposes (e.g. walk to math or reading groups)

mathematical understanding). The design also includes the designated goals or standards that the assessment is aligned to (I.e. CCCS or TEKs) and the district-adopted standards for practice. Design also includes whether the test is adaptive (does it increase or decrease in complexity based on student responses) and whether or not it is scripted for the proctor.	language that is used (which can also be considered a part of the design; hence the overlap with teacher-facing materials and student-facing materials.	words per minute etc.) Results are often reported out in some form and we include these "score reports" as a part of our definition of results when working with TAP. Results can be delivered electronically or in hardcopy.	Actions, for the purpose of our research, also include decisions made that impact the student learning and assessment experience. These actions may be determined externally (E.g.: district or state mandated) or internally (E.g.: grade-level team or classroom-based.)

Appendix B – Written Record of Manual Post-it Note Coding from KJ Technique

Color Key for Double-Coded Data Points	
Assessment Design	3
Implementation	4
Results	7
Instructional Actions	2

Assessment Design	**Implementation**	**Results**	**Instructional Actions**
Appropriate vocabulary and language for age of students Not "trick" ?s or "trick" info Appropriate vocabulary Standard vocabulary Multiple ?s on the same skill Assesses skills in different ways Multiple assessment of the same skill Appropriate language (i.e. English or Chinese) Consider format (multiple choice, short answer, etc.) Differentiated (oral, auditory, computer)	Easy to record Easy to score Not too long Quick to administer Easy for teacher to analyze Short Engaging **Administer (with 8 x votes)**	Starting points Check points Communicate how parents can help at home Use for parent communication To communicate information to parents Color coded reports Hyperlinked reports Reports with intuitive names Communicate with next teacher effectively Goal setting Conferencing Makes sense to share with the student	Designing work Moves instruction forward Create lessons for each group Deficit skills link to appropriate lesson for reteach Guides instruction Grouping To create groups To create groups **Classroom instruction (with 8 x votes)** **Grouping (with 5 x votes)** **Next steps (with 1 x vote)**

Easy for kids to understand Clear instructions Easy to understand Ability to test high and low kids -keep going for those who know more Not give data on reading of a math assessment Identifies needs Identify missing skills Pinpoint exact gaps Shows gaps in learning Shows next steps for that child Creating work based on the need ?s actually align to TEK Aligns vertically Focus on TEKS Content versus STAAR readiness Align with report card Captures student thinking Clearly tests one concept/skill **Alignment (with 7 x votes)**		Helps teacher give feedback while conferencing Objective not subjective Meaningful scale **Feedback (with 1 x vote)** **Teacher facing info (with 1 x vote)**	

Vocab (with 3 x votes)			