journal of mixed methods studies

Issue 10, 2024
Journal of Mixed Methods Studies / JOMES

Just When You Thought that Quantitizing Merely Involved Counting: A Renewed Call for Expanding the Practice of Quantitizing in Mixed Methods Research With a Focus on Measurement-Based Quantitizing

Anthony J. Onwuegbuzie[1]

To cite this article:

Onwuegbuzie, A. J. (2024). Just when you thought that quantitizing merely involved counting: A renewed call for expanding the practice of quantitizing in mixed methods research with a focus on measurement-based quantitizing. *Journal of Mixed Method Studies*, *10,* 99-148. www.jomesonline.com, https://doi.org/10.59455/jomes.54

Article Info: *Received*: 16 July 2024 *Revised*: 2 September 2024 *Accepted*: 9 October 2024

Abstract
In this article, I explore the concept of quantitizing in mixed methods research, categorizing it into four types: *descriptive-based quantitizing* (i.e., converting qualitative data into quantitative summaries; e.g., frequencies), *inferential-based quantitizing* (i.e., using statistical methods to draw inferences from quantitized data), *exploratory-based quantitizing* (i.e., identifying patterns/relationships within quantitized data, often leading to further quantitative analysis), and *measurement-based quantitizing* (i.e., applying psychometric models to quantitized data to assess and to measure latent traits). Among these, measurement-based quantitizing is the least prevalent. Therefore, I expand the concept of measurement-based quantitizing by demonstrating how modern test theory (MTT) approaches (e.g., Rasch analysis and item response theory [IRT] models) can be applied effectively to quantitized themes or finer data units like categories, codes, and sub-codes. Rasch analysis and foundational IRT models (1-parameter IRT, 2-parameter IRT, 3-parameter IRT) add significant value to descriptive-based quantitizing by providing deeper insights into theme difficulty and discrimination. Other IRT models (e.g., 4-parameter IRT, 5-parameter IRT, Bayesian IRT) offer further refinement. Also, I highlight the value of these models in inferential-based quantitizing, particularly via differential item functioning analysis. When applying IRT to quantitized themes, tools such as the test information function, item characteristic curves, and item fit analysis are essential for refining measurements. I underscore the importance of optimizing theme quantity and sample size, recommending minimum guidelines for reliable IRT analysis of quantitized themes. In conclusion, I call for the broader adoption of measurement-based quantitizing, integrating MTT approaches to enhance the rigor, precision, and interpretative power of mixed methods research.

Keywords: mixed methods research, quantitizing, descriptive-based quantitizing, inferential-based quantitizing, measurement-based quantitizing, exploratory-based quantitizing, DIME-model of quantitizing, modern test theory, latent traits, Rasch analysis, item response theory, psychometrics, theme analysis, differential item functioning, theme optimization, qualitative data analysis, quantitative analysis, 1 + 1 = 1 integration, full(er) integration, integrated mixed methods research

[1] **Corresponding author**: University of Johannesburg, Johannesburg, South Africa; University of South Africa, Pretoria, South Africa; https://orcid.org/0000-0002-4569-57962

Just When You Thought that Quantitizing Merely Involved Counting: A Renewed Call for Expanding the Practice of Quantitizing in Mixed Methods Research With a Focus on Measurement-Based Quantitizing

The field of mixed methods research has advanced significantly since Teddlie and Tashakkori stated in 2003 that "the [mixed methods] field is just entering its 'adolescence' and that there are many unresolved issues to address before a more matured mixed methods research area can emerge" (p. 3). These rapid developments led Onwuegbuzie and Hitchcock, in 2019, to declare that the mixed methods research field has entered "emerging adulthood" (p. 18).

Interestingly, one of the biggest developments has been the development of what Onwuegbuzie and Combs (2010) coined as "cross-over mixed analyses" (p. 422)—hereafter referred to as *crossover mixed analyses*—wherein one or more analysis types associated with one tradition (e.g., qualitative analysis) are used to analyze data associated with a different tradition (e.g., quantitative data). As concluded by Teddlie and Tashakkori (2009), "We believe that this [use of crossover mixed analyses] is one of the more fruitful areas for the further development of MM [mixed methods] analytical strategies" (p. 281). According to Onwuegbuzie and Combs (2010), crossover analyses can be used to reduce, to display, to transform, to correlate, to consolidate, to compare, to integrate, to assert, or to import data—yielding nine crossover analysis types, which are presented in Table 1.

Table 1

Crossover Mixed Analysis Strategies

Analysis Step	Cross-Case Analysis Strategy
Integrated Data Reduction	Reducing the dimensionality of qualitative data/findings using quantitative analysis (e.g., exploratory factor analysis of qualitative data) and/or quantitative data/findings using qualitative techniques (e.g., thematic analysis of quantitative data) (Onwuegbuzie, 2003, 2024; Onwuegbuzie & Teddlie, 2003)
Integrated Data Display	Visually presenting both qualitative and quantitative results within the same display (Lee & Greene, 2007; Onwuegbuzie, 2024; Onwuegbuzie & Dickinson, 2008)
Data Transformation	Converting quantitative data into data that can be analyzed qualitatively (i.e., *qualitizing* data; Onwuegbuzie & Leech, 2019; Tashakkori & Teddlie, 1998), and/or qualitative data into numerical codes that can be analyzed statistically (i.e., *quantitizing* data; Miles & Huberman, 1994; Onwuegbuzie, 2024; Samdelowski et al., 2009; Tashakkori & Teddlie, 1998)
Data Correlation	Correlating qualitative data with quantitized data and/or quantitative data with qualitized data (Onwuegbuzie, 2024; Onwuegbuzie & Teddlie, 2003)
Data Consolidation	Combining or merging multiple data sets to create new or consolidated codes, variables, or data sets (Louis, 1982; Onwuegbuzie, 2024; Onwuegbuzie & Teddlie, 2003)
Data Comparison	Comparing qualitative and quantitative data/findings (Onwuegbuzie, 2024; Onwuegbuzie & Teddlie, 2003)

Analysis Step	Cross-Case Analysis Strategy
Data Integration	Integrating qualitative and quantitative data/findings either into a coherent whole or two separate sets (i.e., qualitative and quantitative) of coherent wholes (McConney, Rudd, & Ayres, 2002; Onwuegbuzie & Teddlie, 2003)
Warranted Assertion Analysis	Reviewing all qualitative and quantitative data to yield meta-inferences (M. L. Smith, 1997)
Data Importation	Utilizing follow-up findings from qualitative analysis to inform the quantitative analysis (e.g., qualitative contrasting case analysis, qualitative residual analysis, qualitative follow-up interaction analysis, and qualitative internal replication analysis; Li et al., 2000; Onwuegbuzie & Teddlie, 2003) or follow-up findings from quantitative analysis to inform the qualitative analysis (e.g., quantitative extreme case analysis, quantitative negative case analysis; Onwuegbuzie & Teddlie, 2003)

Adapted from "Emergent Data Analysis Techniques in Mixed Methods Research: A Synthesis," by A. J. Onwuegbuzie and J. P. Combs, 2010, *Handbook of mixed methods in social and behavioral research,* p. 422. Copyright 2010 by Sage.

Among these nine types of crossover analyses, the only type that does not necessitate both qualitative and quantitative data to be collected within the same study is *data transformation*. In general, transformation involves the conversion of one data type into another to facilitate combined analysis. As noted in Table 1, in their simplest forms, data transformation involves *qualitizing* and/or *quantitizing*.

The technique of qualitizing involves transforming quantitative data into a qualitative form that can be analyzed qualitatively (Onwuegbuzie & Leech, 2019; Tashakkori & Teddlie, 1998). An effective way of qualitizing data is by constructing narrative profiles wherein narrative descriptions are constructed from quantitative data (Tashakkori & Teddlie, 1998). There are several types of narrative profile formation, including the following traditional profiles: modal profiles, holistic profiles, average profiles, comparative profiles, and normative profiles.

- **Modal Profiles**: These profiles are detailed narrative descriptions of a group of individuals of interest that are formed based on the most frequently occurring characteristics of the group that they represent.
- **Holistic Profiles**: These profiles represent the researcher's overall impression of the unit of investigation.
- **Average Profiles**: These profiles are derived from the mean of several attributes of individuals or situations.
- **Comparative Profiles**: These profiles are created by comparing one unit of analysis to one or more other units, highlighting their similarities and differences.
- **Normative Profiles**: These profiles, akin to narrative profiles, are based on comparing an individual or group to a standard, such as a normative group (cf. Onwuegbuzie & Teddlie, 2003; Tashakkori & Teddlie, 1998).

In addition to these traditional narrative profiles, there are at least additional 38 narrative profiles that can be used to qualitatize data, as follows: exemplar profiles, dynamic profiles, cluster profiles, contextual profiles, thematic profiles, trajectory profiles, persona profiles, intersectional profiles, critical profiles, transformative profiles, narrative arc profiles, comparative contextual profiles, sentiment profiles, aggregated profiles, behavioral profiles, predictive profiles, interactional profiles, conflict profiles, latent profiles, cultural profiles, evolutionary profiles, causal profiles, aspirational profiles, identity profiles, spatial profiles, marginalized profiles, decision-making profiles, network profiles, experiential profiles, historical profiles, relational profiles, value-based profiles, event-centric profiles, context-sensitive profiles, developmental profiles, resilience profiles, ethical or moral profiles, and geographic profiles. Each of these profiles is described as follows:

- **Exemplar Profiles**: These profiles highlight specific cases or individuals who exemplify a particular characteristic or behavior within the group being studied. Exemplar profiles are useful for illustrating extreme or particularly illustrative examples within the data set.
- **Dynamic Profiles**: These profiles capture changes over time in an individual's or group's attitudes, behaviors, or experiences. They are particularly useful in longitudinal studies wherein the goal is to track development or shifts in perceptions or actions.
- **Cluster Profiles**: Cluster profiles are constructed by grouping individuals or cases who share similar characteristics or experiences. These profiles are used to identify and to describe subgroups within the data, allowing for the exploration of diversity and variation within the larger group.
- **Contextual Profiles**: These profiles integrate the influence of situational or environmental factors on individuals or groups. They emphasize how context shapes behaviors, attitudes, or outcomes and particularly are useful in ecological or contextual analyses.
- **Thematic Profiles**: Derived from recurring themes in the data, these profiles focus on capturing the essence of qualitative findings centered on thematic groupings rather than on individual or aggregate characteristics.
- **Trajectory Profiles**: Similar to dynamic profiles but with a specific focus on the pathway or sequence of events leading to an outcome. These profiles often are used in process-oriented studies like life histories or career trajectories.
- **Persona Profiles**: These are often used in user research and human-centered design to represent archetypical users based on shared goals, behaviors, and attitudes. They synthesize qualitative and quantitative data into fictional yet representative profiles.
- **Intersectional Profiles**: These profiles analyze the interaction between multiple identity factors (e.g., race, gender, socioeconomic status) and how these intersections influence experiences or outcomes.
- **Critical Profiles**: Focused on deconstructing power dynamics, ideologies, or systemic structures affecting the unit of analysis. These profiles are rooted in critical theory and aim to uncover hidden assumptions or societal influences.
- **Transformative Profiles**: Highlight the impact of interventions or changes on individuals or groups, documenting shifts in perspectives, behaviors, or outcomes. These are particularly relevant in evaluative research or action-oriented studies.
- **Narrative Arc Profiles**: Constructed to represent the "storyline" of a group or individual, these profiles incorporate beginning, middle, and end phases, emphasizing the sequence and resolution of key events or processes.
- **Comparative Contextual Profiles**: A hybrid of comparative and contextual profiles, focusing on comparing units of analysis across different environmental or situational contexts to highlight the role of external influences.
- **Sentiment Profiles**: Built from emotional or attitudinal data, these profiles provide a nuanced look at the affective dimensions of individuals or groups, often using qualitative descriptions derived from sentiment analysis.
- **Aggregated Profiles**: Synthesizing diverse data sources into a single narrative to provide a comprehensive overview. These profiles are commonly used in meta-syntheses or integrative reviews.
- **Behavioral Profiles**: Focused on patterns of action or decision-making within the unit of analysis, offering insights into behavioral tendencies or routines.
- **Predictive Profiles**: Leveraging existing data to construct profiles that forecast future behaviors, decisions, or outcomes, often integrating qualitative interpretations with statistical trends.

- **Interactional Profiles**: These profiles emphasize the dynamics and relationships between individuals or groups, highlighting interactions and their effects on collective outcomes or perceptions.
- **Conflict Profiles**: Designed to explore tensions, disagreements, or conflicts within or between groups, these profiles focus on understanding sources, dynamics, and resolutions of conflict.
- **Latent Profiles**: Derived from hidden patterns or underlying variables within data, these profiles focus on factors not immediately apparent but significant in shaping the narrative (e.g., unconscious biases or implicit attitudes).
- **Cultural Profiles**: Highlighting cultural influences and shared norms within a group, these profiles emphasize the role of cultural identity in shaping behaviors, attitudes, or outcomes.
- **Evolutionary Profiles**: Focused on tracing the development or transformation of a phenomenon over extended periods, these profiles emphasize slow or generational changes in attributes or behaviors.
- **Causal Profiles**: Explicitly constructed to investigate cause-and-effect relationships within the data, focusing on how specific variables or interventions impact individuals or groups.
- **Aspirational Profiles**: These profiles focus on the goals, hopes, or aspirations of individuals or groups, emphasizing potential or desired outcomes.
- **Identity Profiles**: Constructed to explore how personal or collective identities are formed, expressed, and negotiated within the group of interest.
- **Spatial Profiles**: Integrating geographical or locational data into narrative descriptions, these profiles explore the role of physical space or environment in shaping the unit of analysis.
- **Marginalized Profiles**: Highlighting the experiences of individuals or groups at the margins of societal norms or systems, these profiles focus on voices often underrepresented in traditional analyses.
- **Decision-Making Profiles**: Exploring how individuals or groups arrive at decisions, these profiles emphasize cognitive processes, contextual factors, and outcomes.
- **Network Profiles**: Constructed by mapping connections and relationships within a group, these profiles focus on social or organizational networks and their influence on behavior or information flow.
- **Experiential Profiles**: Focus on detailed accounts of lived experiences, highlighting the subjective perspectives of individuals or groups. These profiles often are used in phenomenological studies.
- **Historical Profiles**: Constructed using historical data to explore how past events or trends influence current behaviors or outcomes. These profiles are valuable in historical or sociological research.
- **Relational Profiles**: Focus on relationships and interactions between individuals or groups, emphasizing dynamics like power, collaboration, or conflict. Useful in studies of social networks or organizational behavior.
- **Value-Based Profiles**: Explore the core values, beliefs, or principles guiding individuals or groups, often constructed from qualitative data like interviews or focus groups.
- **Event-Centric Profiles**: Constructed around specific events or phenomena, these profiles analyze how a significant occurrence impacts individuals or groups, offering a detailed narrative of its ripple effects.
- **Context-Sensitive Profiles**: These profiles take into account variability in individual or group responses based on shifting contexts or environments, emphasizing adaptability or situational dependence.
- **Developmental Profiles**: Highlight stages of growth or change within individuals or groups, focusing on progression or regression across a timeline, often in education, psychology, or child development research.
- **Resilience Profiles**: Focus on coping mechanisms, adaptive strategies, and recovery pathways for individuals or groups facing challenges or adversity.
- **Ethical or Moral Profiles**: Explore ethical frameworks, decision-making processes, or moral reasoning within the unit of analysis, often constructed from qualitative insights in bioethics or moral psychology.

- **Geographic Profiles**: Integrate spatial data to construct narratives that emphasize location-based differences or patterns, often useful in geographic or environmental studies.

Recently, Onwuegbuzie and Leech (2019) expanded the definition of qualitizing wherein qualitizing can involve five major elements (see also Onwuegbuzie & Leech, 2021). Specifically, these authors conceptualized that qualitizing

1. can yield numerous representations,
2. can stem not only from quantitative data but also from qualitative data,
3. can involve qualitative analyses and/or quantitative analyses,
4. can involve a single analysis or multiple analyses, and
5. can yield a fully integrated analysis.

As a result, qualitizing has the potential to facilitate multimodal narratives, to democratize evidence, and to leverage research in the post-truth era wherein the authority of researchers can be ignored or even undermined, and wherein misinformation easily can spread. Therefore, Onwuegbuzie and Leech (2019, 2021) contend that the inclusion of qualitizing techniques helps to validate and to contextualize quantitative findings, ensuring that the findings are both accurate and contextually rich. Further, they demonstrate that qualitizing can lead to qualitative insights that add layers of meaning to quantitative findings, thereby mitigating the impact of false narratives by providing a more nuanced understanding of data. Notwithstanding, the authors advocate for the continued expansion of fully integrated qualitizing approaches to enhance research outcomes, emphasizing the need for more robust methodologies and tools that can balance effectively the qualitative-quantitative divide.

In contrast, as stated in Table 1, in its simplest form, *quantitizing* involves the process of converting qualitative data into numerical codes that can be analyzed statistically (Miles & Huberman, 1994; Onwuegbuzie, 2024; Onwuegbuzie & Teddlie, 2003; Sandelowski et al., 2009; Tashakkori & Teddlie, 1998). Although within the social sciences, discussion of the concept of data transformation, in general (Caracelli & Greene, 1993), and coining of the term *quantitizing* and its variants (i.e., *quantitize*, *quantitized*, *quantitizes, quantitising*, *quantitise*, *quantitised*, *quantitises*), in particular (Tashakkori & Teddlie, 1998), can be traced back more than a quarter of a century to the 1990s, until recently, the concept of quantitizing has been framed as inherently involving some form of counting or tallying occurrences, frequencies, or instances of specific qualitative elements. Moreover, the concept of quantitizing historically has lacked an organizing framework. However, recently, Onwuegbuzie (2024) conceptualized a comprehensive meta-framework that encompasses a broader range of quantitizing techniques beyond mere counting, involving complex judgments and multiple analytical techniques to convert qualitative data into quantitative forms. This meta-framework categorizes quantitizing at three levels. Most notably, at the first level—the foundational level—quantitizing into four distinct types: **D**escriptive-based quantitizing, **I**nferential-based quantitizing, **M**easurement-based quantitizing, and **E**xploratory-based quantitizing. The first letters of these four types form the acronym "DIME," which led to this meta-framework being called the *DIME-Model of Quantitizing*. As noted by Onwuegbuzie (2024), this DIME acronym functions not only as a mnemonic, but also a metaphor for the utility and precision inherent in the meta-framework that is represented by these techniques.

Specifically, descriptive-based quantitizing involves converting qualitative data into quantitative metrics to describe the characteristics of the data (Onwuegbuzie, 2024). This approach involves the application of statistical techniques to summarize and to convey the patterns found in qualitative responses. It includes classes of descriptive statistics such as *measures of central tendency* (i.e., identifying the central point or typical value in a qualitative data set; e.g., mean, median, mode), *measures of variation/dispersion* (i.e., describing the spread or variability within a qualitative data set; e.g., range, interquartile range, variance, standard deviation), *measures of position/relative standing* (i.e., indicating the relative position of specific data points within a qualitative data set; percentiles, quartiles, z scores, t scores), and *measures of distributional shape* (i.e., describing the shape and nature of the distribution of qualitative data; e.g., skewness, kurtosis). In general, descriptive-based quantitizing involves employing statistical methods to generate a detailed summary of qualitative data, enabling the structured and quantifiable representation of the data's core elements and subtleties. This approach is particularly advantageous during the initial phases of data analysis, whereby identifying overall trends and patterns is essential (Onwuegbuzie, 2024).

Inferential-based quantitizing involves transforming qualitative data into estimations or predictions, allowing researchers to extend conclusions beyond the immediate data set to broader contexts (Onwuegbuzie, 2024). Techniques include general linear model (GLM) analysis such as correlation coefficients, independent/dependent samples *t* tests, analysis of variance (ANOVA), multiple regression analysis, canonical correlation analysis, structural equation modeling (SEM), and hierarchical linear modeling (HLM). The overall aim of inferential-based quantitizing is to draw inferences about a broader population from a sample, thereby allowing researchers to extend their conclusions beyond the specific qualitative data set to wider contexts. This method is crucial for validating hypotheses and supporting generalizations. Essentially, inferential-based quantitizing involves converting qualitative insights into quantitative data that can be statistically analyzed, thereby facilitating conclusions about larger populations and offering a robust tool for expanding research findings in mixed methods research studies (Onwuegbuzie, 2024).

Measurement-based quantitizing involves transforming qualitative data into quantitative data specifically for the purposes of measurement (Onwuegbuzie, 2024). These purposes include the development of instruments (focusing on ensuring the score reliability and score validity of measurement tools) and score-validation of constructs. Analytical techniques that are used for measurement-based quantitizing include Rasch modeling and item response theory (IRT). In essence, measurement-based quantitizing involves systematically converting qualitative observations into quantifiable scales or metrics, enabling empirical testing and score-validation. It facilitates the measurement of complex constructs that, while not directly observable, can be inferred from qualitative data (Onwuegbuzie, 2024).

Lastly, exploratory-based quantitizing is used to uncover new insights and patterns within qualitative data, employing methodologies such as exploratory factor analysis, principal components analysis, cluster analysis, or correspondence analysis (Onwuegbuzie, 2024). It is particularly useful in the early stages of research for generating new hypotheses and theories based on the data. For example, after conducting in-depth interviews to collect qualitative data, exploratory-based quantitizing could be used to convert these interview responses into quantitative data. These quantitative data then could be subjected to a cluster analysis to identify distinct groups among the interview participants. These groupings then can reveal insights that can inform the refinement of future qualitative, quantitative, and/or mixed methods research studies (Onwuegbuzie, 2024).

Conclusions Regarding the DIME-Model of Quantitizing

As can be seen, the DIME-model provides a structured approach to quantitizing, categorizing it into four distinct types: Descriptive, Inferential, Measurement, and Exploratory. Each type serves a unique purpose, allowing mixed methods researchers to select the most appropriate level of quantitizing-based analysis according to their research questions and the nature of their data (Onwuegbuzie, 2021). By providing this comprehensive framework, the DIME model enhances the depth and rigor of mixed methods research studies, facilitating a more integrated and thorough analysis of qualitative data, thereby enriching the analytical process and broadening the scope of research findings. It enables a systematic and nuanced transformation of qualitative data into quantitative forms, allowing for a more comprehensive analysis that integrates both qualitative richness and quantitative precision. This structured approach not only broadens the scope of research findings, but also strengthens the overall methodological robustness of studies, making it a valuable tool in mixed methods research.

Furthermore, the DIME-model's structured approach supports the dynamic needs of mixed methods research by offering flexibility and precision. Moreover, it stands out as a versatile and comprehensive framework that significantly enhances the analytical capabilities of mixed methods researchers. In so doing, it fosters a more holistic understanding of complex and complicated research questions, ultimately leading to richer, more informed, and more impactful research outcomes. As such, the DIME-model has much potential for researchers aiming to leverage the strengths of both qualitative and quantitative data in their mixed methods research studies (Onwuegbuzie, 2024).

Prevalence of the Four Types of Quantitizing

Most Common Quantitizing Technique: Descriptive-Based Quantitizing

Of the four types of quantitizing conceptualized by Onwuegbuzie (2024), the technique of descriptive-based quantitizing—considered the lowest form of quantitizing because it involves the use of Level 1 analysis, the lowest form of statistical analysis (cf. Figure 1)—has been utilized by mixed methods researchers far more than have the other forms. In fact, the vast majority of quantitizing that is undertaken by mixed methods researchers involves some form of descriptive-based quantitizing (Ross & Onwuegbuzie, 2014). In particular, qualitative data—taking the form of codes, categories, sub-themes, themes, figures of speech, meta-themes, narratives (i.e., prose or poetry), and the like—have been subjected to descriptive-based quantitizing via the computation of mean prevalence rates. For instance, Onwuegbuzie et al. (2007) investigated the perceptions of 912 university students regarding the characteristics of effective college teachers. These students completed a questionnaire in which they were asked to identify and to rank between three and six characteristics that they believed effective instructors should possess or demonstrate, and to provide a definition or description for each characteristic. A qualitative analysis, specifically using constant comparison analysis, of these responses led to the identification of the following nine themes that were summarized by the acronym RESPECTED: **R**esponsive, **E**nthusiast, **S**tudent-centered, **P**rofessional, **E**xpert, **C**onnector, **T**ransmitter, **E**thical, and **D**irector. Onwuegbuzie et al. (2007) quantitized these themes in the following way:

> if a student listed a characteristic that was eventually unitized under a particular theme, then a score of 1 would be given to the theme for the student response; a score of 0 would be given otherwise. This dichotomization led to the formation of an interrespondent matrix (i.e., Student × Theme Matrix) (Onwuegbuzie, 2003; Onwuegbuzie & Teddlie, 2003). Both matrices consisted only of 0s and 1s. (p. 127)

The researchers then used this inter-respondent matrix to undertake descriptive-based quantitizing, specifically, "By calculating the frequency of each theme from the inter-respondent matrix, percentages were computed to determine the prevalence rate of each theme" (p. 127), which led to the numerous insightful findings, such as that student-centeredness was the most prevalent theme (i.e., mean frequency = 58.88%) (see Table 2).

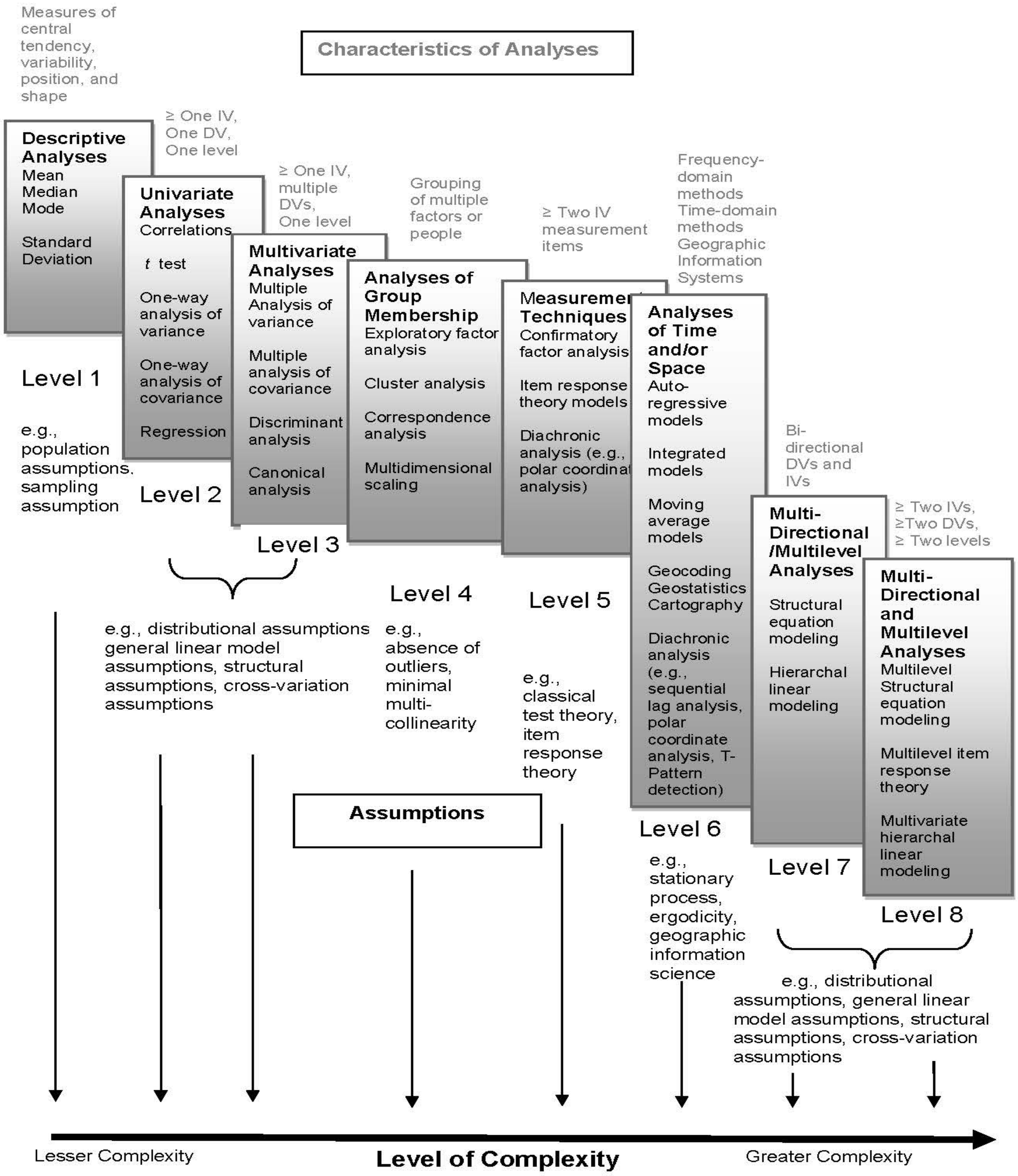
That descriptive-based quantitizing is the most commonly used quantitizing technique has intuitive appeal, bearing in mind that descriptive data analysis often serves as a bridge between exploratory data analysis and confirmatory data analysis (Abt, 1987). This technique allows researchers to summarize and to convey general trends and patterns within the qualitative data, providing a foundation for further statistical analyses. By converting qualitative data into descriptive statistics, researchers easily can identify key themes and insights, which then can be explored in greater depth via one or more of the other three quantitizing techniques (i.e., inferential-based quantitizing, measurement-based quantitizing, and exploratory-based quantitizing).

Second Most Common Quantitizing Technique: Inferential-Based Quantitizing

Although significantly less utilized than is descriptive-based quantitizing, inferential-based quantitizing is the next most common technique. At its most integrated level, this involves correlating quantitized themes with characteristics of study participants who generated those themes. These associations represent what Onwuegbuzie (2017) referred to as the *1 + 1 = 1 integration approach* (see also Hitchcock & Onwuegbuzie, 2022; Natesan et al., 2019; Newman et al., 2015; Onwuegbuzie, 2023; Onwuegbuzie & Hitchcock, 2019, 2022; Onwuegbuzie et al., 2018). This approach is a meta-framework

Figure 1

Quantitative Analysis Complexity Continuum



Adapted from "Complexity of Quantitative Analyses Used in Mixed Research Articles," by A. Ross and A. J. Onwuegbuzie, 2014. *International Journal of Multiple Research Approaches, 8*, p. 66. Copyright 2014 by Dialectical Publishing

Table 2

Themes Emerging from Students' Perceptions of the Characteristics of Effective College Instructors

Theme	Endorsement Rate (%)
Student-Centered	58.88
Expert	44.08
Professional	40.79
Enthusiast	29.82
Transmitter	23.46
Connector	23.25
Director	21.82
Ethical	21.60
Responsive	5.04

Adapted from "Students' perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed methods analysis," by A. J. Onwuegbuzie, A. E. Witcher, K. M. T. Collins, J. D. Filer, C. D. Wiedmaier, & C. W. Moore, 2007. *American Educational Research Journal, 44*, p. 132. Copyright 2007 by Sage.

that involves "the optimal mixing, combining, blending, amalgamating, incorporating, joining, linking, merging, consolidating, or unifying of research approaches, methodologies, philosophies, methods, techniques, concepts, language, modes, disciplines, fields, and/or teams within a single study" (Onwuegbuzie & Hitchcock, 2022, p. 598). In particular, this approach represents the full integration of quantitative and qualitative components at the data collection, data analysis, and data interpretation stages of the mixed methods research process. However, unfortunately, the overwhelming majority of studies utilizing inferential-based quantitizing only involves the lowest levels of inferential analyses—specifically, univariate analyses such as correlation coefficient, independent/dependent samples *t* test, and analysis of variance—which represent only Level 2 analyses in Figure 1. However, there are a few instances wherein a higher level of inferential-based quantitizing takes place. For instance, McClure et al. (2021) performed a Level 3 analysis (cf. Figure 1) as part of the quantitizing process by conducting a canonical correlation analysis to explore the multivariate relationship between the three most prevalent (quantitized) themes (i.e., Living Environment, Technology-Related, and Overall Learning Experience) that represented the challenges face by students from a university in New York City that hindered your ability to learn successfully online during the lockdowns as a result of the COVID-19 pandemic and six socio-demographic variables (i.e., *gender* [men vs. women]; *age group* [18-37 vs. ≥ 38]; *race* [White vs. Non-White]; *full-time status* [full-time vs. part-time]; *level of student* [undergraduate vs. graduate]; and *technology access* [full access to a computer and reliable Internet vs. not full access to a computer and/or non-reliable Internet]). As can be seen from Table 3, the canonical correlation analysis led to the following finding:

> Overall, the selected canonical correlation solution, which was statistically significant (F[8, 196] = 2.52, p = .013) and practically significant (Canonical R_{c1} = .42; moderate effect size), indicated that the multivariate relationship was mostly characterized by the relationship between age, race, full-time status, and technology access on the socio-demographic side and Technology-Related and Overall Learning Experience on the challenge themes side. (p. 157)

Table 3

Canonical Solution for First Function: Relationship Between the Three Demographic Variables and the Three of the Four Most Prevalent Themes[5]

Variable	Standardized Coefficient	Structure Coefficient	$Structure^2$ (%)
Demographic:			
Age group (18-37 vs. 38+)	-.48*	-.71*	50.4
Race (White vs. Non-White)	.46*	.30*	10.9
Status (Full-time vs. Part-time)	-.45*	-.73*	53.3
Technology Access (full access vs non-full access)	-.39*	-.50*	25.0
Theme:			
Overall Learning Experience	.85*	.95*	90.3
Technology-Related	.34*	.57*	32.5

*Coefficients with the effect sizes larger than .3 (Lambert & Durand, 1975).

Adapted from "Online learning challenges experienced by university students in the New York City area during the COVID-19 pandemic: A mixed methods study," by D. R. McClure, E. O. Ojo, M. B. Schaefer, D. Bell, S. S. Abrams, & A. J. Onwuegbuzie, 2021. *International Journal of Multiple Research Approaches, 13*(2), p. 158. Copyright 2021 by Dialectical Publishing.

This example illustrates how inferential-based quantitizing—in this case, via the use of canonical correlation analysis (see also Onwuegbuzie, 2022)—can lead to meaningful insights that might not be as easily discerned through purely qualitative or descriptive-based quantitized techniques, highlighting its value in comprehensive research endeavors. In particular, inferential-based quantitizing allows researchers to explore complex, multivariate relationships among qualitative variables or between qualitative variables and quantitative variables. In turn, this helps researchers to understand how different qualitative variables might interact with each other in a holistic context, providing a deeper insight than do themes alone. Furthermore, this technique can help to identify which qualitative variables predict other variables. Such findings can be crucial for developing targeted interventions or policies based on the predictive power of certain qualitative variables over others. In addition, inferential-based techniques facilitate researchers in making full use of the qualitative data collected, turning qualitative data into quantifiable trends that can be statistically analyzed. This transformation allows researchers to apply a range of statistical tools and techniques that can uncover subtler trends and patterns not visible through qualitative analysis alone. Also, by employing statistical analyses, researchers can test the generalizability of their findings across different populations or settings. This is particularly important in fields wherein researchers often are looking to apply findings broadly.

As such, it is important that mixed methods researchers not only conduct Level 4 inferential analyses during the quantitizing process, but also undertake higher level inferential analyses, especially those at Level 7 (i.e., multi-directional/multilevel analyses involving bi-directional dependent variables and independent variables; e.g., structural equation model, hierarchical linear modeling) and at Level 8 (i.e., multi-directional/multilevel analyses involving multi-directional dependent variables and independent variables; e.g., multilevel structural equation model, multivariate hierarchical linear modeling). Indeed, Level 7 inferential analyses allow for the examination of complex relationships, including bidirectional effects wherein themes can serve as both predictors and outcomes. This is useful for understanding feedback loops and reciprocal causation. In particular, structural equation modeling (SEM) enables the inclusion of latent variables that can account for measurement error and unobserved constructs, leading to more accurate estimates. In contrast, hierarchical linear modeling (HLM) is designed to handle nested data structures (e.g., students within schools), accounting for the non-independence of observations and appropriately partitioning variance at different levels. Both SEM and HLM can model interactions between/among themes at different levels (e.g., individual and group levels), helping to understand how thematic relationships might vary across contexts or groups. By modeling dependencies and accounting for nested structures, these approaches can provide more accurate standard errors and test statistics, leading to more reliable inferences. With respect to Level 8 inferential analyses, multilevel structural equation modeling (MSEM) and multivariate hierarchical linear modeling (MHLM) allow for the simultaneous modeling of multiple dependent variables, which is useful when studying phenomena that involve several outcomes that might be interrelated (e.g., different dimensions of themes). These analyses also are capable of handling multilevel data structures, extending the benefits of HLM and SEM to scenarios involving multiple levels of analysis (e.g., individuals, groups, organizations). Furthermore, they can model the interdependencies between/among multiple outcomes, helping to capture the full complexity of relationships in the qualitative data, including correlations and causal pathways between outcomes. In addition, these models allow for the investigation of cross-level effects, wherein predictors at one level (e.g., demographic variables) influence outcomes at another level (e.g., individual thematic outcomes). And with the ability to model complex interrelationships and dependencies, researchers can test more sophisticated hypotheses, such as mediating or moderating effects, within a multilevel context. Also, both categories of analyses provide richer insights into the qualitative data by allowing for the consideration of multiple pathways and levels of influence; offer great flexibility in handling complex data structures, accommodating various types of variables and relationships; and by accounting for dependencies and hierarchies in the data, produce more robust and reliable estimates. Overall, multi-directional and multilevel analyses are powerful tools for uncovering nuanced relationships in qualitative data, especially when dealing with complex, nested, or multi-dimensional data structures. They help researchers to gain a more comprehensive understanding of the phenomena under study, making them invaluable in many fields, including the social sciences, science, technology, medicine, and health.

Third Most Common Quantitizing Technique: Exploratory-Based Quantitizing

Exploratory-based quantitizing is the third most common technique. This technique primarily involves the identification of group memberships wherein groups can be animals, minerals, or vegetables. More specifically, in the context of social science research, health research, and similar fields, participants, constructs, variables, and items, in particular, can be grouped. In the extant literature, these elements typically have been subjected to exploratory-based quantitizing via exploratory factor analysis, principal components analysis, cluster analysis, or correspondence analysis—all of which represent Level 4 analysis (see Figure 1). For example, circling back to Onwuegbuzie et al.'s (2007) study, these researchers employed a principal component analysis (PCA) on the nine RESPECTED themes. Prior to this, they converted the zero-order correlation coefficients into tetrachoric correlation coefficients to suit the binary data (i.e., "1" vs. "0") more appropriately. This approach was pivotal for tailoring the PCA to the specific nature of the data. Their analysis revealed the following four meta-themes: **c**ommunicator (43.7% of characteristics per meta-theme), **a**dvocate (81.0%), **r**esponsible (41.1%), and **e**mpowering (59.6%). These meta-themes collectively were abbreviated as CARE. Integrating the CARE model with the nine RESPECTED themes resulted in the creation of the *CARE-RESPECTED Model of Effective College Teaching*.

This PCA example provides an illustration of how exploratory-based quantitizing of qualitative data can be used to group *emergent themes* after they have been quantitized. As noted previously, exploratory-based quantitizing of qualitative data can be used to group *participants*—whether the study involves a relatively large sample or a relatively small sample. An example of exploratory-based quantitizing involving grouping a large sample is that of Onwuegbuzie et al. (2020), who examined the challenges experienced by 1,932 students at Stellenbosch University that hindered their ability successfully to learn online during the emergency remote teaching necessitated by the COVID-19 pandemic. The analysis of the open-ended responses using the topic modeling function of WordStat 8.0.29 (Provalis Research, 2020) resulted in the identification of the following seven themes that categorized the students' challenges: Internet Connection, Mental Health, Personal Challenges/Ability, Time Management, Easily Distracted, Family Members/Make it Difficult; and Lecturers.

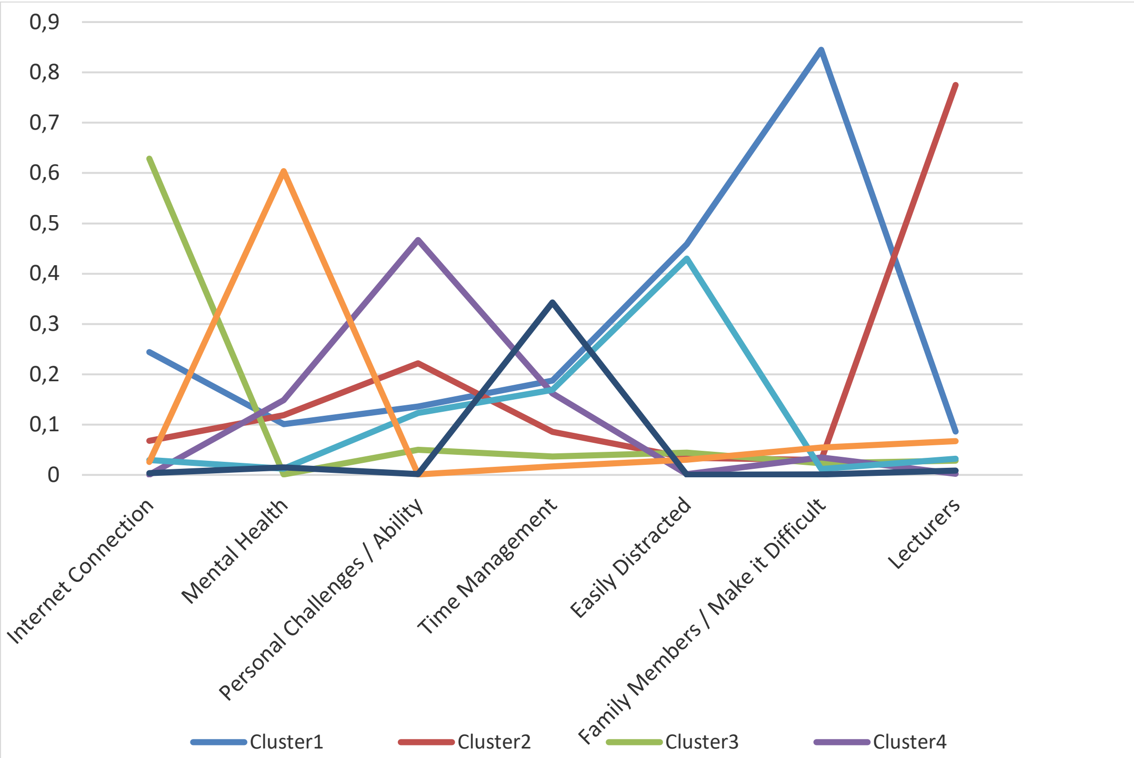
Among the array of analyses conducted by these researchers, they performed a latent class analysis to identify the smallest number of clusters (i.e., latent classes) that could explain the relationships among the seven emergent themes. This analysis was conducted under the assumption that the student participants could be grouped into a few distinct clusters, known as latent classes, based on their profiles related to these emergent themes. The latent class analysis of the seven emergent themes led to the identification of a seven-cluster solution (L^2 = 73.42, df = 72, p = .43, Bootstrap p = .17). These seven distinct clusters of student participants, which are displayed in Figure 2, contained the following proportion of students: Cluster 1 (25.74%), Cluster 2 (19.74%), Cluster 3 (14.84%), Cluster 4 (14.17%), Cluster 5 (10.61%), Cluster 6 (9.08%), and Cluster 7 (5.82%). As can be seen in Figure 2, the profiles of the seven clusters were reported by the researchers as follows:

- "Cluster 1: Students belonging to this cluster are relatively low in terms of the Internet Connection, Mental Health, Personal Challenges/Ability, Time Management, and Lecturers themes; moderate with respect to the Easily Distracted theme; and relatively high in terms of the Family Members/Make it Difficult theme. Therefore, this cluster was labeled by the researchers as students who represent the *Easily Distracted and the Family Members/Make it Difficult cluster*.
- Cluster 2: Students belonging to this cluster are relatively low in terms of all themes except the Lecturers theme, wherein virtually all (i.e., 98%) of participants in this cluster reported experiences that were classified under this theme. Therefore, this cluster was labeled as students who represent the *Lecturers cluster*.
- Cluster 3: Students belonging to this cluster are relatively low in terms of all themes except the Internet Connection theme, wherein virtual all (i.e., 97%) of participants in this cluster reported experiences that were classified under this theme. Therefore, this cluster was labeled as students who represent the *Internet Connection cluster*.
- Cluster 4: Students belonging to this cluster are relatively low in terms of all themes except the Personal Challenges/Ability theme, wherein virtual all (i.e., 98%) of participants in this cluster reported experiences that were classified under this theme. Therefore, this cluster was labeled as students who represent the *Personal Challenges/Ability cluster*.
- Cluster 5: Students belonging to this cluster are relatively low in terms of all themes except the Easily Distracted theme, wherein virtually all (i.e., 95%) of participants in this cluster reported experiences that were classified under this theme. Therefore, this cluster was labeled as students who represent the *Easily Distracted cluster*.
- Cluster 6: Students belonging to this cluster are relatively low in terms of all themes except the Mental Health theme, wherein virtually all (i.e., 92%) of participants in this cluster reported experiences that were classified under this theme. Therefore, this cluster was labeled as students who represent the *Mental Health cluster*.
- Cluster 7: Students belonging to this cluster are relatively low in terms of all themes except the Time Management theme, wherein virtually all (i.e., 95%) of participants in this cluster reported

experiences that were classified under this theme. Therefore, this cluster was labeled as students who represent the *Time Management cluster*." (p. 268)

As can be seen, the latent class analysis conducted in Onwuegbuzie et al.'s (2020) study is a prime example of the utility of exploratory-based quantitizing. The detailed profiles of the clusters allow for the development of targeted interventions and support strategies. For instance, students in the "Mental Health" cluster might benefit from different support services compared to those in the "Internet Connection" cluster. This analysis allows researchers to draw on the strengths of both qualitative and quantitative traditions, providing a more comprehensive understanding of the research problem.

Figure 2
Profiles of Students With Respect to the Seven Emergent Themes



Adapted from "*Challenges experienced by students at Stellenbosch University that hinder their ability successfully to learn online during the COVID-19 era: A demographic and spatial analysis*" by A. J. Onwuegbuzie, E. O. Ojo, A. Burger, T. Crowley, S. P. Adams, & B. T. Bergsteedt, 2020, *International Journal of Multiple Research Approaches*, *12*(3), p. 269. Copyright 2020 by Dialectical Publishing.

Providing an example of how exploratory-based quantitizing qualitative data can be used to group a small number of participants, Onwuegbuzie et al. (2011) conducted a study whose primary purpose was to compare and to contrast pedagogical approaches used by first-generation instructors of mixed methods research courses. To this end, they interviewed, via semi-structured interviews, eight first-generation instructors (4 women, 4 men) of mixed methods research courses from institutions representing various geographic areas in the United States. A constant comparison analysis (Glaser, 1965) led to the emergence of the following three metathemes that represented dimensions of teaching style of the eight mixed methods research instructors: (a) Orientation (i.e., Methodological Focused vs. Question/Topic Focused), (b) Level of Application (i.e., Conceptual vs. Applied), and (c) Level of Structure (i.e., Structured vs. Exploration).

Each of these metathemes was quantitized via a matrix of co-occurrences of themes within each metatheme, which yielded three matrices. Next, each matrix was analyzed using correspondence analysis, which is an exploratory multivariate technique that factors categorical variables—in this case, the metathemes—and maps them in a property space to display their associations in two or more dimensions (Michailidis, 2007). Consequently, the three correspondence analyses produced three correspondence plots, one of which is shown in Figure 2. This figure demonstrates how Onwuegbuzie et al. (2011) utilized QDA Miner, a computer-assisted qualitative data analysis software, to map the eight participants onto a two-dimensional grid using the Orientation metatheme, which comprised two themes and six subthemes. From Figure 2, it is evident that four instructors (Instructors 1, 3, 4, and 6) exhibited a methodology-focused pedagogy, emphasizing a broad approach to scientific inquiry with general preferences for certain types of designs, sampling logic, and analytical strategies. The other four instructors (Instructors 2, 5, 7, and 8) displayed a pedagogy focused on questions and topics. Among the methodology-focused instructors, three (Instructors 3, 4, and 6) employed defined quantitative and qualitative concepts, method-driven research, and knowledge of specific research traditions, suggesting less specificity in methodological orientation. Similarly, three of the question/topic-focused instructors (Instructors 5, 7, and 8) leaned towards a broad research tradition, indicating less specificity in their question/topic orientation. The remaining two question/topic-focused instructors demonstrated higher specificity by providing very detailed research question-driven and topic-driven examples. Onwuegbuzie et al. (2011) also noted that “However, each of the instructors presented her or his pedagogical practices in a unique manner such that the thrust of their pedagogical orientations is uniquely depicted on the continuum” (pp. 182-183).

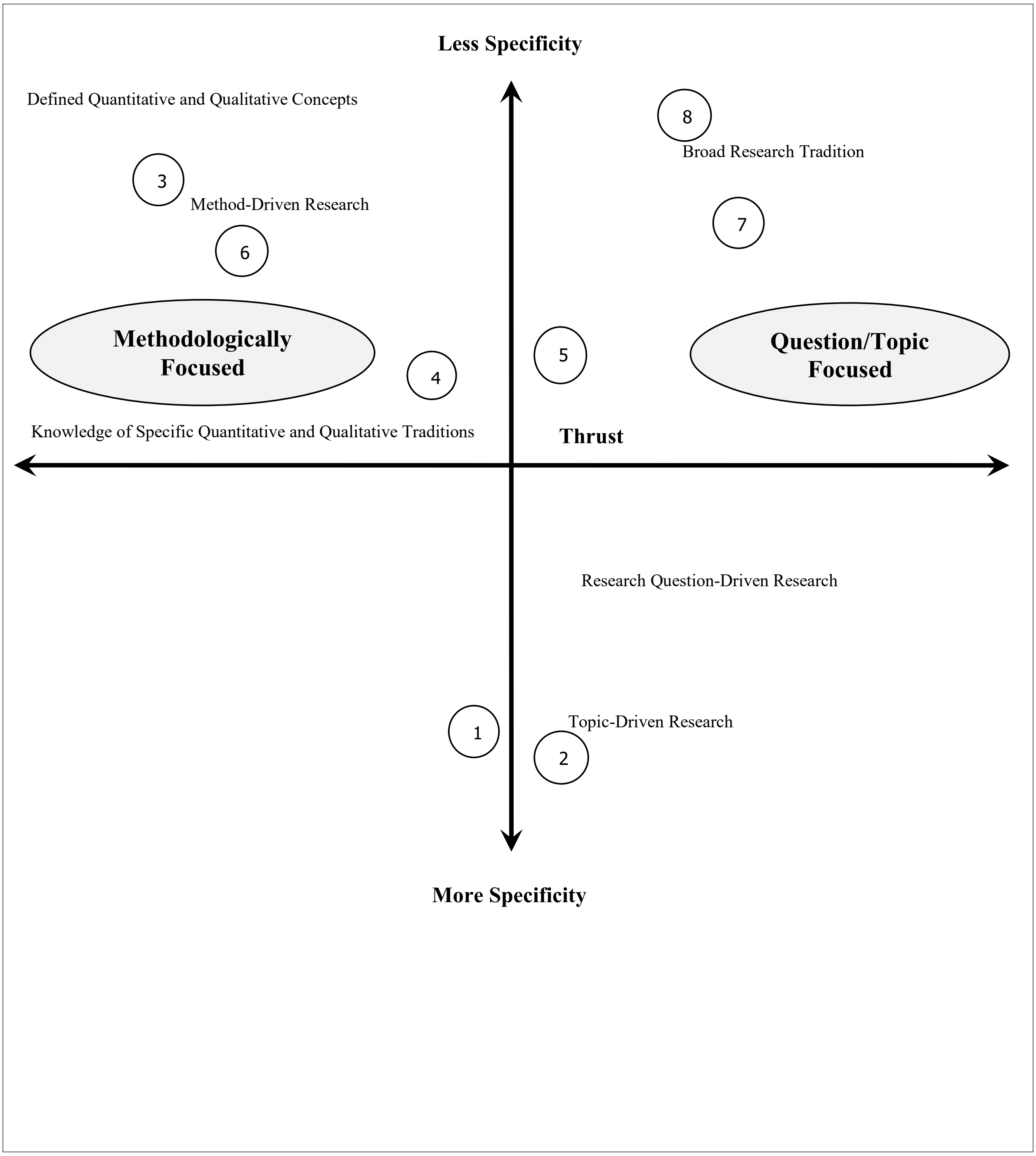
Onwuegbuzie et al. (2011) then compared and contrasted the three correspondence plots, which led to the development of modal profiles, which are illustrated in Figure 3. They termed these modal profiles *pedagogical profiles*, aligning with the concept of a partially ordered meta-matrix as described by Miles and Huberman (1994). This meta-matrix showcased the pedagogical profiles of each participant, categorized by their levels of orientation, application, and structure. Five distinct pedagogical profiles emerged from the analysis. For instance, three participants were identified with the Question/Topic, Applied, and Structured profile, abbreviated as the QAS profile, as indicated in the first three rows of Figure 3—wherein, they

> focused on teaching students how to conduct research studies by allowing the question/topic area drive the method in contrast to allowing a specific research tradition (e.g., quantitative or qualitative) to drive the direction of the study; taught mixed research courses whereby students collected and analyzed real data and wrote up their mixed research reports; and taught mixed research courses with maximal structure, whereby students were introduced to models, typologies, and frameworks, and were expected to use these conceptualizations to inform their understandings of mixed research concepts and applications. (Onwuegbuzie et al., 2011, p. 193)

Onwuegbuzie et al. (2011) detailed narratives for each of the five pedagogical profiles, using acronyms for identification. The narrative describing the five-profile typology was notably rich and informative, especially because this study was the first to employ a formal interview process and mixed methods research approaches to examine the context of mixed methods research courses. This depth of narrative surpassed what would have been achieved merely by reporting codes, categories, subthemes, themes, figures of speech, and/or metathemes. Interestingly, since the publication of their study, the authors’ five-profile typology has been validated by numerous instructors of mixed methods research courses. These instructors confirmed that their teaching styles align with one of the five identified pedagogical approaches.

Figure 3

Metatheme of Orientation: Methodological Focused Pedagogy Versus Research Focused Pedagogy



Adapted from "*A mixed research study of pedagogical approaches and student learning in doctoral-level mixed research courses*." by A. J. Onwuegbuzie, R. K. Frels, N. L. Leech, and K. M. T. Collins, 2011, *International Journal of Multiple Research Approaches*, *5*, p. 182. Copyright 2011 by Dialectical Publishing.

Figure 4

Partially Ordered Meta-Matrix: Pedagogical Profiles of Each of the Participants as a Function of Level of Orientation, Level of Application, and Level of Structure

Participant	Orientation	Application	Structure	Total in Group
2	Question/Topic	Applied	Structured	
5	Question/Topic	Applied	Structured	
7	Question/Topic	Applied	Structured	3
1	Methodological	Conceptual	Exploratory	
6	Methodological	Conceptual	Exploratory	2
8	Question/Topic	Conceptual	Exploratory	1
3	Methodological	Conceptual	Structured	1
4	Methodological	Applied	Structured	1

Adapted from "*A mixed research study of pedagogical approaches and student learning in doctoral-level mixed research courses*." by A. J. Onwuegbuzie, R. K. Frels, N. L. Leech, and K. M. T. Collins, 2011, *International Journal of Multiple Research Approaches*, *5*, p. 193. Copyright 2011 by Dialectical Publishing.

As can be seen, exploratory-based quantitizing serves as a robust methodology in mixed methods research, allowing researchers to apply quantitative techniques to qualitative data. In addition to exploratory factor analysis, principal components analysis, cluster analysis, or correspondence analysis, other analyses that fall into the category of exploratory-based quantitizing include the following:

- **Independent Component Analysis (ICA)**: Similar to principal component analysis, but focuses on making the components statistically independent rather than just uncorrelated (Tharwat, 2021). It involves decomposing a multivariate signal into additive, independent non-Gaussian signals, effectively grouping the variables based on their independence.
- **Biclustering (or Co-Clustering)**: Simultaneously clusters rows and columns of a matrix to find submatrices of the original data matrix that exhibit coherent patterns (Madeira & Oliveira, 2004). It involves assigning both variables and observations into subgroups wherein similar rows and similar columns are grouped together.
- **T-SNE (t-Distributed Stochastic Neighbor Embedding)**: A dimensionality reduction technique particularly well-suited for visualizing high-dimensional datasets by reducing them to two or three dimensions (Wattenberg et al., 2016). It involves projecting high-dimensional data into a lower-dimensional space, often revealing clusters of similar observations.
- **Self-Organizing Maps (SOM)**: An unsupervised learning method that reduces dimensions and groups data using neural networks (Cottrell et al., 2018). It involves organizing data into a grid of nodes wherein similar data points are mapped to the same or neighboring nodes, effectively creating clusters. Further, it involves assigning objects to nodes on the map, effectively grouping similar objects together.
- **Latent Dirichlet Allocation (LDA)**: Primarily used in topic modeling to discover the underlying topics in a collection of documents (Blei et al., 2003). It involves grouping words into topics, wherein each topic represents a latent group of related words.
- **Hierarchical Clustering**: Builds a hierarchy of clusters by either progressively merging or splitting existing clusters (Murtagh & Legendre, 2014). It involves producing a dendrogram,

wherein objects are grouped into clusters based on their similarity at various levels of the hierarchy.

- **Multiple Correspondence Analysis (MCA)**: Extends correspondence analysis to handle more than two categorical variables (Le Roux & Rouanet, 2010). It involves mapping variables into a low-dimensional space to identify patterns and relationships among multiple categorical variables.
- **Non-negative Matrix Factorization (NMF)**: Decomposes a matrix into the product of two lower-rank matrices with non-negative elements, often used for pattern recognition (Lee & Seung, 2001). It involves grouping variables and observations based on their contribution to the factorized matrices.
- **Hierarchical Cluster Analysis**: To group objects (e.g., themes, variables, individuals) into a tree-like structure of nested clusters (Murtagh, 1983). It involves assigning objects to clusters based on their similarity, forming a hierarchy of groups.
- **K-means Clustering**: To partition objects (e.g., themes, variables, individuals) into a specified number of clusters based on their characteristics (Celebi et al., 2013). It involves assigning objects to clusters by minimizing the variance within each cluster.
- **Multifactor Dimensionality Reduction (MDR)**: To detect and to characterize interactions among variables, particularly in genetic studies (Hahn et al., 2003). It involves identifying combinations of variables (or factors) that best explain the variance in the data.
- **Multidimensional Scaling (MDS)**: To visualize the similarity or dissimilarity of data in a low-dimensional space (Hout et al., 2013). It involves mapping objects into a spatial configuration wherein similar objects are closer together, effectively grouping them.
- **Affinity Propagation**: To identify exemplars among data points and form clusters based on similarity (Bombatkar & Parvat, 2015). It involves assigning objects to clusters with the most representative data points as exemplars.
- **Latent Semantic Analysis (LSA)**: To analyze relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms (Reidy, 2009). It involves grouping terms and documents into latent semantic structures.
- **Kohonen Networks (Self-Organizing Maps)**: To project high-dimensional data onto a lower-dimensional grid while preserving the topological properties of the data (Ultsch, 1990). It involves grouping data points based on their location on the map, with similar points mapped to nearby nodes.
- **Factor Mixture Modeling (FMM)**: To identify latent classes and factors simultaneously, combining aspects of factor analysis and mixture modeling (Lubke & Muthén, 2005). It involves grouping observations into latent classes, and variables are associated with factors within those classes.
- **Latent Profile Analysis (LPA)**: Similar to latent class analysis but used for continuous observed variables (Spurk et al., 2020). It involves assigning individuals to latent profiles based on their continuous data, identifying distinct subgroups within the population.
- **Gaussian Mixture Models (GMMs)**: A probabilistic model for representing the presence of subpopulations within an overall population without requiring that the subpopulations be labeled (Viroli & McLachlan, 2019). It involves assigning observations to clusters based on the probability of belonging to each cluster.
- **Agglomerative Hierarchical Clustering**: A bottom-up clustering method that builds a hierarchy of clusters (Bombatkar & Parvat, 2015; Bouguettaya et al., 2015). It involves grouping observations iteratively together based on their similarities, forming a dendrogram that illustrates the nested grouping.

These approaches facilitate the identification of new patterns and relationships, providing valuable insights that can shape future research and can inform decision-making processes. By integrating qualitative and quantitative research techniques, researchers enhance their analytical toolkit, enabling a more thorough and nuanced exploration of their data. This integrative methodology bridges the gap between qualitative and quantitative research traditions, fostering a holistic understanding of complex phenomena. Moreover, exploratory-based quantitizing unveils subtle nuances and intricate dynamics within the data that might otherwise remain obscured. This method offers a richer, more detailed landscape for academic inquiry, enabling researchers to capture the complexity of real-world issues more effectively. As a result, it not only strengthens theoretical frameworks, but also enhances practical applications, offering actionable insights that can drive improvements in various fields. In summary, the power of exploratory-based quantitizing lies in its ability to transform qualitative insights into quantifiable data, enriching the research process and leading to a more comprehensive understanding of the subject matter. This technique enables the discovery of hidden patterns, detailed profiling, statistical validation, and targeted interventions. It enhances the depth and breadth of research findings, making it a valuable tool for researchers aiming to integrate qualitative richness with quantitative precision. Moreover, exploratory-based quantitizing empowers researchers to uncover deeper insights and to develop more informed strategies, ultimately advancing both academic and practical knowledge.

Least Common Quantitizing Technique: Measurement-Based Quantitizing

As noted previously, measurement-based quantitizing primarily involves the use of Rasch modeling and item response theory (IRT). Rasch modeling is a specific form of IRT that applies a one-parameter logistic model to the analysis of test data (Bond & Fox, 2020; Wright & Masters, 1982). It emphasizes the creation of measurements that are invariant across different groups of respondents and items (Wright & Masters, 1982). The Rasch model ensures that the probability of a correct response is a logistic function of the difference between the person's ability and the item difficulty. It is used extensively in educational testing, health outcomes measurement, and psychology for developing and score-validating assessments. The utility of the Rasch model includes the fact that it provides interval-level measurement from ordinal data, it ensures that item parameters (difficulty) and person parameters (ability) are estimated independently, and it allows for the creation of scales wherein the measurement properties of items are invariant across different subgroups.

IRT is a family of models used to analyze the relationship between latent traits (i.e., unobserved characteristics) and their manifestations (e.g., observed responses on tests or questionnaires) (Hambleton et al., 1991; Lord, 1980). It includes one-parameter models (e.g., Rasch; Andrich, 2004; Fischer, 1995), two-parameter models (which also estimate discrimination parameters; Hambleton et al., 1991; Lord, 1980), three-parameter models (which additionally estimate guessing parameters; Barton & Lord, 1981; Hambleton et al., 1991; Lord, 1980), four-parameter models (which additionally estimate of the *upper asymptote* or *D-parameter*, which accounts for the maximum probability that even the most knowledgeable examinee might incorrectly answer an item; Kalkan, 2022), and five-parameter models (which include an asymmetry parameter; Gottschalk & Dunn, 2005). IRT is used widely in educational testing (e.g., standardized tests), psychological assessments, and health outcomes research. The utility of IRT include that it allows for more complex modeling of items and responses, it can handle a wide range of item types and response formats, and it supports computerized adaptive testing (CAT) wherein the test adapts to the test-taker's ability level in real-time.

Both Rasch modeling and IRT are central to modern test theory, which focuses on the development and refinement of tests and assessments (McDonald, 2013). The aim of modern test theory is to improve the precision and fairness of measurement tools, ensuring that they yield reliable and valid scores across different populations and contexts (McDonald, 2013). The major difference between Rasch modeling and IRT is that Rasch modeling, a specific type of IRT, is considered to be more stringent in its requirements but simpler in its assumptions and applications. In contrast, IRT is a broader framework that includes models with more parameters, offering greater flexibility for complex assessments and adaptive testing environments.

Prevalence of Measurement-Based Quantitizing Driven by Modern Test Theory

Of the four types of quantitizing conceptualized by Onwuegbuzie (2024), the technique of measurement-based quantitizing—considered a moderate to high form of quantitizing because it can involve the use of analysis from Level 5 to as high as Level 8 (cf. Figure 1)—has been utilized by mixed methods researchers far less than has the other forms of quantitizing. As evidence of this claim, of the 364 articles (i.e., not including editorials, media reviews, notes, errata) published in the *Journal of*

Mixed Methods Research (*JMMR*) at the time of writing (i.e., from Volume 1 and Issue 1 to Volume 18 and Issue 3), only 6 articles mentioned Rasch analysis in either the abstract or body (i.e., David et al., 2018; Koskey et al., 2018; Koskey & Stewart, 2014; Kramer, 2011; Leach Sankofa, 2022; Morell & Tan, 2009). This frequency indicates that 1.65% (i.e., 6 / 364) of all published *JMMR* articles to date have discussed/utilized Rasch analysis to some degree.

Specifically, Morell and Tan (2009) applied Rasch analysis using the partial credit model to assess the internal structure and psychometric properties of a science assessment for fifth- and sixth-grade students. Their Rasch analysis provided item difficulty estimates, person ability estimates, and item/person fit statistics, which helped to validate the hierarchical structure of the assessment. Think-aloud interviews and exit interviews were conducted to collect qualitative data. The goal of these interviews was to understand the cognitive processes of students while they responded to the assessment items. The student surveys collected data on perceived item difficulty and fairness, which then were coded and categorized for analysis. The Rasch analysis identified items with misfit statistics or unexpected threshold gaps, indicating potential issues with certain items. The qualitative data from think-aloud and exit interviews provided context and explanations for the Rasch analysis findings, highlighting why specific items were difficult or easy for students. For instance, items identified as problematic by Rasch analysis were investigated further via qualitative data, leading to revisions that addressed specific student misunderstandings or misinterpretations. The integration of qualitative data ensured that the assessment items not only were statistically valid, but also were meaningful and interpretable from the students' perspectives. This comprehensive approach of combining Rasch analysis with qualitative data collection methods, such as think-aloud and exit interviews, allowed for a robust validation process that addressed both quantitative and qualitative aspects of assessment validity.

Kramer (2011) conducted Rasch analysis, specifically the Rasch partial credit model and the mixed Rasch model, to assess the psychometric properties of the Child Occupational Self-Assessment (COSA). This Rasch model was used to estimate item difficulty and person ability, as well as to examine the fit of items and respondents within the COSA. Their study involved the integration of Rasch analysis results with qualitative data from cognitive interviews, semi-structured interviews, and observations. Qualitative data were used to understand how children interpreted and responded to COSA items, providing context to the Rasch analysis findings. Moreover, the qualitative data helped to identify items that did not fit well in the Rasch model by revealing the reasons behind respondents' difficulties or misunderstandings with specific items. Cognitive interviews and observations provided insights into how children perceived the COSA items, informing item revisions and ensuring that the items were both psychometrically sound and meaningful from the children's perspectives. The integration of qualitative data allowed for a more nuanced understanding of the social validity of the COSA, addressing the acceptability and relevance of the assessment to the children who used it. The study demonstrates how qualitative data can validate and explain quantitative findings from Rasch analysis, leading to a more comprehensive assessment validation process. By combining quantitative Rasch analysis with qualitative insights, the researcher was able to triangulate data sources to provide a robust evaluation of the COSA's social validity, ensuring that the assessment was both reliable and relevant to its intended users.

Koskey and Stewart (2014) conducted Rasch analysis using the many-facet Rasch model to assess the psychometric properties of Absolute Magnitude Estimation Scales (MES) used in the Quantitative Attitudes Questionnaire (QAQ). Their Rasch model was used to transform the MES data into interval measures and to evaluate the fit of items and respondents within the scales. Their study involved the integration of qualitative data from think-aloud protocols and follow-up interviews with quantitative results from Rasch analysis. Respondents completed a think-aloud task while rating items using MES, followed by interviews about their scale use and definitions. The Rasch analysis results led to the identification of disordered categories and step calibrations in the MES data, indicating issues with the scale's monotonicity and meaningfulness. The qualitative data provided context for these findings, revealing how respondents interpreted and used the scales, highlighting discrepancies in their cognitive processes. The think-aloud protocols and interviews explained why certain items did not fit the Rasch model well, particularly highlighting that respondents often did not use MES in a meaningful, consistent way. Moreover, qualitative insights informed the interpretation of Rasch analysis results by demonstrating that respondents' definitions and use of MES were inconsistent, affecting the quantitative meaningfulness of the data. The authors concluded that the integration of qualitative and quantitative data was essential for understanding the cognitive processes behind respondents' use of MES, leading to recommendations for improving scale instructions and ensuring better alignment between respondent intentions and scale use. By combining Rasch analysis with qualitative methods, the study provided a comprehensive evaluation of the MES's effectiveness and meaningfulness, demonstrating the importance of mixed methods in survey research.

David et al. (2018) conducted Rasch analysis—specifically, the Rasch Rating Scale model—to develop and to score validate an instrument measuring trust among college patients receiving athletic training services. Their Rasch analysis focused on item fit, item difficulty, and the score reliability of the instrument. Fit statistics such as infit mean squares and outfit mean squares were used to evaluate how well the items fit the Rasch model and to identify items that needed revision. Qualitative data were collected via cognitive interviews and semi-structured interviews with college patients and athletic trainers. The purpose of these interviews was to understand how respondents interpreted and responded to the survey items, providing insights into the context and nuances of the trust construct. The study also involved the use of a table of specifications developed from qualitative themes to guide item writing and to ensure that all relevant aspects of the trust construct were covered. The initial qualitative inquiry provided a foundational understanding of the trust construct, which informed the item development process. Emergent themes from the interviews guided the creation of survey items that were contextually relevant and meaningful to the respondents. Cognitive interviews were conducted iteratively between rounds of Rasch analysis to refine and to improve the items. These interviews helped to identify issues with item wording, respondent interpretation, and the appropriateness of response options. Results from the Rasch analysis, such as items with poor fit statistics, were examined using qualitative data to understand the reasons behind misfits. This iterative process of qualitative feedback and quantitative validation ensured that the final instrument was both psychometrically sound and contextually valid. Specific examples include the revision of items related to communication and accessibility based on feedback from cognitive interviews, which then were re-evaluated using Rasch analysis to confirm improvements in fit and reliability. By combining Rasch modeling with qualitative methods, the researchers were able to develop a robust instrument for measuring trust, demonstrating the value of a mixed methods research approach in instrument development. The qualitative data provided essential context and insights that informed the Rasch analysis, leading to a more nuanced and comprehensive understanding of the trust construct and a more valid and reliable measurement tool. This article highlights the iterative and integrative process of using both qualitative and quantitative research methods to enhance the development and score-validation of research instruments, ensuring that they are both statistically rigorous and meaningful to the target population.

Koskey et al. (2018) used Rasch analysis to evaluate the psychometric properties of the Transformative Experience Questionnaire (TEQ). This analysis was applied to examine item difficulty, person ability, and the fit of each item and respondent to the model. It helped to determine the functioning of the TEQ's rating scale, ensuring that the scale categories were used as intended and that each category represented a distinct progression along the measured variable. The researchers integrated qualitative data from think-aloud protocols and cognitive interviews. These qualitative methods were used to understand how students interpreted and responded to the TEQ items. Qualitative data provided insights into students' thought processes, revealing how they understood the items and the response categories. Rasch analysis led to the identification of items that did not fit well with the model, indicating potential issues with item wording or conceptual clarity. These misfitting items then were examined further using qualitative data from cognitive interviews to understand the reasons behind the misfit. For example, cognitive interview data revealed that some students interpreted the phrase "in my everyday life" to include school-related activities, leading them to rate themselves lower than intended. This qualitative insight informed the revision of the items to align better with the intended construct. Conversely, the qualitative data helped to identify issues with specific items even when Rasch analysis did not indicate a misfit. For instance, qualitative feedback suggested that certain items were misunderstood due to complex wording, prompting revisions that improved the overall clarity and functionality of the questionnaire. The integration of Rasch analysis and qualitative data allowed for a comprehensive validation process, ensuring that the TEQ was both psychometrically sound and meaningful from the students' perspectives. Overall, this article demonstrates how combining Rasch analysis with qualitative methods can enhance the development and score-validation of educational instruments, ensuring that they yield reliable and valid scores, and are contextually appropriate.

Finally, Leach Sankofa (2022) developed and exemplified the Transformativist Measurement Development Methodology (TMDM) through the creation of the Peer Bonds Scale. This researcher aimed to enhance the rigor of scale construction methodologies by integrating social justice principles, particularly within marginalized communities. The methodology involved six stages: checking assumptions, setting parameters, inductively operationalizing the construct, qualitatively generating items, quantitatively examining psychometric properties, and examining trustworthiness. Leach Sankofa (2022) conducted Rasch analysis as part of mixed methods strategies in scale development. Specifically, the author acknowledged Rasch modeling as representing one of the frameworks used to demonstrate various approaches to scale construction. This article situated Rasch modeling within the broader context of creating score reliable and score valid constructs in mixed methods research.
According to the author, in the TMDM, qualitative data are crucial in the early stages of construct development. These stages include the qualitative generation of items and the inductive operationalization of the construct. The qualitative findings inform the psychometric analysis conducted subsequently, ensuring that the generated scale items are rooted in the lived experiences and perspectives of the target population. The integration of qualitative data into the Rasch analysis framework ensured that the scale items were both culturally sensitive and contextually relevant. Qualitative insights shaped the initial item pool, which then was refined via quantitative methods like Rasch analysis to assess item validity and reliability. This bidirectional process enhanced the rigor and trustworthiness of the scale, ensuring it accurately reflected the construct that it aimed to measure. The TMDM exemplified a robust integration of qualitative and quantitative research methods, including Rasch analysis, to develop scales that were both psychometrically sound and socially just. This approach elevates participant voices and aligns measurement practices with transformative social science principles.

Item Response Theory

With respect to item response theory (IRT), of the same *JMMR* 364 articles, only 4 articles mentioned IRT in either the abstract or body. This frequency indicates that 1.10% (i.e., 4 / 364) of all published *JMMR* articles to date have discussed/utilized IRT to some degree. These four articles comprise three articles described earlier that also conducted a Rasch analysis (i.e., David et al., 2018; Leach Sankofa, 2022; Morell & Tan, 2009) and a fourth article by Onwuegbuzie et al. (2010).

Benefits of Using Rasch Modeling and Item Response Theory Within a Mixed Methods Research Framework

These *JMMR* articles provide compelling evidence that combining Rasch modeling and/or IRT—that is, modern test theory—with qualitative methods allows for a thorough validation process addressing both statistical rigor and contextual relevance. This integration ensures that the instruments developed yield reliable and valid scores across different contexts and populations. Qualitative data provide critical context and explanations for the quantitative findings from modern test theory analyses, helping to identify the underlying reasons for item misfits or unexpected response patterns, leading to more informed revisions and improvements.

The iterative process of qualitative feedback and quantitative validation via modern test theory ensures continuous refinement of measurement instruments, addressing issues as they arise and leading to more accurate and meaningful assessments. Triangulating data from multiple sources enhances the robustness of research findings, allowing researchers to provide a more comprehensive evaluation of the instruments and constructs being measured. This mixed methods research approach not only enhances the statistical rigor of the instruments, but also ensures that they are meaningful and interpretable for the target populations.

Some studies have demonstrated that integrating qualitative data with modern test theory approaches ensures that the developed instruments are culturally sensitive and socially relevant. This alignment with the lived experiences and perspectives of the target population enhances the acceptability and applicability of the assessments. Incorporating modern test theory within a mixed methods research framework offers a holistic approach to measurement development, capturing the complexity of constructs and ensuring that instruments are both psychometrically sound and meaningful to respondents.

In summary, the use of modern test theory within a mixed methods research framework across these articles demonstrates its strengths in creating robust and contextually relevant measurement instruments. The integration of qualitative data enriches the validation process, leading to more nuanced and comprehensive understandings of the constructs being measured. This approach exemplifies the power of combining qualitative and modern test theory methods to enhance the development and score-validation of research instruments. As such, I call for more mixed methods researchers and psychometricians to incorporate modern test theory approaches within a mixed methods research framework.

Towards a New Framework for Measurement-Based Quantitizing

All of the aforementioned publications were innovative, powerfully illustrating how modern test theory approaches can be integrated with qualitative methods to achieve better research outcomes. However, it should be noted that none of the seven modern test theory-based articles (i.e., six involving Rasch analysis and one discussing IRT) directly applied Rasch or IRT analysis to *qualitative* data. Instead, these modern test theory-based approaches involved the analysis of *quantitative* data, and then were integrated with qualitative methods to evaluate and to score-validate instruments, combining modern test theory metrics with qualitative insights to inform instrument revisions and to ensure measurement reliability and validity. As such, although not a criticism of these articles, none of them actually involved measurement-based quantitizing. That is, these articles did not involve either the Rasch analysis or IRT analysis of quantitized themes or other types of quantitized data (e.g., codes, categories, subthemes, metathemes). Yet, such modern test theory analyses go well beyond mapping out the theme prevalence that is characterized by descriptive-based quantitizing. Therefore, for the remainder of this article, I will outline how modern test theory approaches that are appropriate for analyses at the level of theme or lower (e.g., categories, codes, sub-codes) can be used for the purpose of measurement-based quantitizing.

Applying Rasch Analysis and Foundational Item Response Theory Models to Quantitized Qualitative Data: A Value-Added Approach to Descriptive-Based Quantitizing

Because Rasch and IRT inherently are designed to focus on item-level analysis—modeling the relationship between latent traits (e.g., ability or proficiency) and individual test items—they offer a robust framework that can be effectively applied to theme-level analysis stemming from qualitative research. This adaptability arises from their ability to quantify and to model the interaction among underlying constructs (e.g., themes or concepts in qualitative data) and lower level data points—that is, data closer to the raw (i.e., original) qualitative data (e.g., categories, codes, and sub-codes derived from qualitative coding). When qualitative data are quantitized, these measurement models allow researchers to assess how well specific themes capture the latent constructs that they are intended to measure, ensuring that the relationships between these themes and the underlying traits are represented systematically and accurately.

Furthermore, Rasch and IRT provide detailed diagnostics at the item level—analogous to the code or sub-code level in qualitative data—enabling researchers to refine their coding schemes and to improve the trustworthiness of their analyses. This item-level focus allows these models to offer precise insights into how each element—whether it is a broad theme or a granular sub-code—contributes to the overall measurement of the underlying construct. By quantitizing qualitative data and applying these models, researchers can assess whether a particular theme consistently reflects a latent trait across different respondents or contexts, much like how Rasch or IRT would analyze whether a test item is appropriately difficult or discriminating in a traditional assessment.

Moreover, these models can be applied to analyze not only broad themes, but also more granular elements of qualitative data, such as categories, codes, and sub-codes. This multi-level capability particularly is valuable in qualitative research, wherein data can be nested and hierarchically structured. By applying Rasch or IRT models to sub-codes, researchers can gain insights into the specific nuances and patterns within their data, identifying which sub-codes are most informative and which might be redundant or less effective in capturing the intended construct.

In summary, the strengths of Rasch and IRT in item-level analysis make them exceptionally well-suited to the intricacies of theme-level analysis in qualitative research. By leveraging these models, researchers can assess quantitatively the relationships between various levels of coded qualitative data and their corresponding latent constructs, leading to more rigorous, trustworthy, and nuanced interpretations of qualitative findings. This approach not only enhances the precision of theme-level analysis, but also supports the development of more sophisticated and validated coding frameworks that can withstand the complexities of real-world qualitative data.

Rasch Analysis and One-Parameter Item Response Theory

There are some conceptual and philosophical differences between the Rasch model and the one-parameter item response theory (1PL IRT) model, particularly in their origins, assumptions, and applications (Andrich, 2004; Fischer, 1995). Both models assume that all items have equal discrimination, but the Rasch model particularly is concerned with ensuring that data fit the model to achieve specific objectivity in measurement. In contrast, the 1PL IRT model often is used as a starting point in a broader set of IRT models, which can be adapted depending on the data and the goals of the analysis. This makes the 1PL IRT model more flexible, whereas the Rasch model remains more rigid in its application, focusing on the construct validity of the measurement scale (Fischer, 1995). However, both models are mathematically equivalent when applied to dichotomous data, which means that they often yield the same results in terms of estimating item (theme) difficulty and participant ability (engagement) (Hayat et al., 2020). Therefore, with respect to their use in measurement-based quantitizing of themes, they can be discussed together.

Both Rasch analysis and 1PL IRT analyses can provide more nuanced insights into the themes identified in qualitative research beyond their mere frequency of occurrence that is obtained via descriptive-based quantitizing. When applied to the inter-respondent matrix, wherein each theme is treated as an item and each participant's contribution to that theme as a response (coded as "1" for contribution and "0" for no contribution), both analyses can yield the following insights:

- **Theme Difficulty:** Both Rasch analysis and 1PL IRT estimate the difficulty of each theme, which, in this context, refers to how likely participants are to contribute to a particular theme. In particular, a theme with high difficulty indicates that fewer participants are coded under that theme, indicating that it is less common. In turn, this suggests that it might be more complex, less relevant, or more challenging for participants with which to engage. Conversely, a low-difficulty theme would be one to which many participants contribute.

- **Participant Ability:** Both models also estimate participant ability, reflecting how likely a participant is to contribute to themes in general. A higher ability score would indicate a participant who is more engaged or who has contributions across many themes. That is, a higher ability score means that a participant is more likely to contribute to a wide range of themes, particularly those with higher difficulty.

- **Item-Person Map:** Both analyses generate an item-person map (i.e., Wright map), which visually represents the distribution of theme difficulties and participant abilities on the same scale. This allows for the identification of which themes are relatively more or less common (i.e., more challenging for participants) and which participants are more or less engaged overall.

- **Measurement Invariance:** Both analyses also can test for measurement invariance across different groups (e.g., gender, age), ensuring that the themes are interpreted similarly across these groups, thereby allowing for meaningful comparisons.

In summary, both approaches allow researchers to understand not just how often themes occur, but also how they vary in difficulty and how participant contributions vary in ability, providing a richer interpretation of the data (Schulz & Fraillon, 2011).

Two-Parameter Item Response Theory

The two-parameter item response theory (2PL IRT) model extends the Rasch model by adding a discrimination parameter (Hambleton et al., 1991; Lord, 1980) for each theme, as follows:

- **Theme Difficulty:** The model still estimates the difficulty of each theme, providing insights into how likely participants are to engage with it.

- **Theme Discrimination:** In addition to difficulty, the 2PL IRT model assesses how well each theme discriminates among participants with different levels of ability. Themes with higher discrimination values are more effective at differentiating between participants who are likely to engage with the theme and those who are not. This helps to identify which themes are most informative about differences in participant engagement.

- **Participant Ability:** Similar to the Rasch model, participant abilities are estimated based on their responses to the themes.

- **Differentiation Among Participants:** The 2PL IRT model provides a more nuanced understanding of participant contributions, particularly by highlighting themes that are more or less effective at distinguishing between high and low contributors.

This model enables a more detailed examination of themes, identifying not only their relative frequency and difficulty, but also how sharply they distinguish among different participants.

Three-Parameter Item Response Theory

In addition to the difficulty and discrimination parameters, the three-parameter item response theory (3PL IRT) model adds a guessing parameter to the analysis (Barton & Lord, 1981; Hambleton et al., 1991; Lord, 1980), which accounts for the possibility that a participant might randomly contribute to a theme:

- **Theme Difficulty and Discrimination:** Like the 2PL IRT model, the 3PL IRT model estimates both the difficulty and discrimination of each theme.

- **Guessing Parameter:** The 3PL IRT model introduces the concept of a "guessing" parameter for each theme. The guessing parameter accounts for the possibility that participants might engage with a theme by chance or for superficial reasons. In the context of themes, this could indicate themes that are more accessible or easier with which to engage even without deep understanding or experience with the underlying phenomenon or construct. More specifically, this parameter reflects the probability that a participant with very low ability (i.e., someone unlikely to contribute meaningfully to many themes) still contributes to a particular theme. In a qualitative context, this might correspond to themes that are so broad or general that even less engaged participants still contribute to them.

- **Participant Ability:** Participant abilities are estimated, taking into account the potential for guessing.

- **Enhanced Understanding of Theme Engagement:** This model allows researchers to identify themes that might be less robust indicators of participant engagement because high guessing parameters could indicate themes that are more likely to attract contributions from less-engaged participants.

By applying the 3PL IRT model, researchers gain insights into which themes are most likely to be contributed to by participants who are otherwise less engaged, providing a more nuanced interpretation of the data.

Four-Parameter Item Response Theory

The four-parameter item response theory (4PL IRT) model includes an additional parameter that includes the asymptote at the upper end of the response curve (i.e., upper asymptote parameter) (Kalkan, 2022), which accounts for the possibility that some themes might not engage fully even the most engaged participants:

- **Theme Difficulty, Discrimination, and Guessing:** The 4PL model includes all the parameters of the 3PL model.

- **Upper Asymptote Parameter:** The upper asymptote parameter reflects the maximum probability that even the most engaged participants will contribute to a theme. This parameter recognizes that even highly engaged participants might not contribute to every theme, perhaps due to the theme's specificity or relevance. As such, this parameter provides insights into themes that are difficult with which to engage fully, even for highly engaged participants.

- **Participant Ability:** Abilities are estimated with consideration for difficulty, discrimination, guessing, and the upper engagement limit.

- **Refined Theme Contribution Analysis:** By accounting for the possibility that some themes might never achieve full contribution from all participants, the 4PL model allows for a more refined analysis of theme engagement, particularly at the upper end of participant engagement.

Therefore, the 4PL IRT model provides a highly detailed view of theme contributions, considering the full range of participant engagement levels and the varying likelihood of theme engagement.

Five-Parameter Item Response Theory

The 5-parameter item response theory (5PL IRT) model includes an asymmetry parameter in addition to the four parameters in the 4PL IRT model (Gottschalk & Dunn, 2005):

- **Theme Difficulty, Discrimination, Guessing, and Upper Asymptote:** The 5PL IRT model estimates all the parameters of the 4PL IRT model.

- **Asymmetry Parameter:** This parameter adjusts for potential skewness in the relationship between participant overall level of engagement and engagement with a theme. It helps to identify themes wherein engagement is not evenly distributed around the difficulty threshold, possibly due to cultural factors or varied interpretations.

- **Participant Ability:** Abilities (i.e., levels of engagement) are estimated considering all five parameters, providing a comprehensive understanding of participant engagement with the themes.

Consequently, the 5PL IRT model offers an exceptionally nuanced analysis of theme contributions, accounting for the entire spectrum of participant engagement and the diverse probabilities of engagement with each theme.

Summary of Rasch Analysis Five Item Response Theory Models

In summary, applying Rasch analysis and IRT models to quantitized qualitative data, such as themes, provides a significantly more detailed and nuanced interpretation of themes than does descriptive-based quantitizing alone. By leveraging these models, researchers can move beyond simply counting the frequency of themes to understanding the underlying patterns in participant engagement. Rasch and 1PL IRT analyses allow for the estimation of theme difficulty and participant ability, offering insights into how challenging certain themes are for participants and how likely participants are to engage with them. These models also produce an item-person map, facilitating the visualization of the relationship between theme difficulty and participant ability on a common scale, and ensuring measurement invariance across different groups.

As we progress to more complex IRT models, additional parameters further enrich the analysis. The 2PL IRT model introduces theme discrimination, allowing researchers to assess how well themes differentiate among participants with varying levels of engagement. The 3PL IRT model adds a guessing parameter, accounting for the likelihood of participants contributing to a theme by chance, which helps identify themes that might attract superficial engagement. The 4PL IRT model includes an upper asymptote parameter, providing insights into themes that might not fully engage even the most engaged participants. Finally, the 5PL IRT model incorporates an asymmetry parameter, allowing for the adjustment of skewness in participant engagement with themes, particularly in culturally or contextually nuanced areas. Overall, these models offer a value-added approach to understanding qualitative data, making the interpretation of themes more sophisticated and comprehensive.

Applying Other Item Response Theory Models to Quantitized Qualitative Data: A Value-Added Approach to Descriptive-Based Quantitizing

In addition to foundational IRT models (i.e., 1PL IRT, 2PL IRT, 3PL IRT, 4PL IRT, 5PL IRT), there are other IRT models that can be applied to quantitized themes. These models include polytomous IRT models, multidimensional IRT models, hierarchical IRT models, mixture IRT models, Bayesian IRT models, and diagnostic classification models (DCMs). Each of these models will be described in the context of analyzing quantitized themes.

Polytomous IRT Models

- **Graded Response Model (GRM):** Graded response models are used for ordinal response data (e.g., Likert scales) (Samejima, 1969). As such, they are ideal for analyzing quantized themes that are ordinal in nature. Themes that are ordinal in nature can result from both *incidence quantitizing*—for example, when the analyst specifies the extent to which the theme was present (e.g., "Strongly Disagree" to "Strongly Agree") and *intensity quantitizing*—for instance, when the analyst specifies how intense the experience underlying theme was (e.g., "not at all intense" to "very intense"). These models estimate both item difficulty (i.e., the points on the engagement continuum at which participants are likely to move from one level of engagement to the next; e.g., difficulty of moving from low to moderate engagement with a theme) and item discrimination (i.e., how well each theme discriminates between participants at different levels of engagement), as well as latent trait scores, which estimate, for each individual, their level of the latent trait (e.g., their general propensity or likelihood to engage with and to contribute to the themes identified in the qualitative data). These models can help in understanding how different levels of agreement with a theme relate to an underlying latent trait. More specifically, graded response models can offer a nuanced understanding of the intensity or level of engagement with each theme by participants.

- **Partial Credit Model (PCM):** Partial credit models are another type of IRT model that also can be used to analyze ordinal data (Masters, 1982), particularly when the response categories reflect increasing levels of engagement. It can provide a detailed analysis of how participants engage with themes at different levels of intensity or complexity. This model estimates the difficulty of each step (i.e., step difficulty) or threshold between categories within a theme. Unlike the graded response mode, partial credit models are not utilized under the assumption of equal discrimination across themes. This approach provides a more nuanced and detailed understanding of engagement, making it possible to identify specific areas where participants struggle and to tailor interventions accordingly.

- **Generalized Partial Credit Model (GPCM):** The generalized partial credit model is a more flexible version of partial credit model, including discrimination parameters (Muraki, 1992). Similar to the partial credit model, the generalized partial credit model estimates step difficulty but unlike the partial credit model, it estimates discrimination parameters, which can vary across themes.

- **Nominal Response Model (NRM):** Nominal response models are a type of IRT model designed to handle categorical data wherein the response options do not have a natural order (De Ayala, 2013). Unlike the graded response model and the partial credit model, which assume an ordinal relationship among responses, nominal response models are ideal for situations wherein the themes are nominal (e.g., different options that represent distinct, non-ordered choices) (Thissen et al., 2011). These models estimate a discrimination parameter (Thissen et al., 2011). For quantitized themes, this means understanding how strongly each specific theme response (e.g., a particular coping strategy) is associated with different levels of participant engagement. Higher discrimination values suggest that the category particularly is effective at distinguishing between participants who are more or less likely to engage with that theme. Furthermore, nominal response models estimate difficulty or location parameters for each response category. In the context of quantitized themes, this can reveal which responses (e.g., strategies or behaviors related to a theme) are more or less likely to be chosen by participants. For example, some coping strategies might be easier or more common, whereas others are more challenging or less frequently endorsed. Also, nominal response models calculate the probability that a participant with a given level of the latent trait will select each response category within a theme. This provides insight into how likely different types of participants (e.g., those with high or low engagement) are to choose each response option. It allows researchers to predict and to understand the distribution of responses across different levels of participant traits.

- **Rating Scale Model (RSM):** The rating scale model is a specialized case of the partial credit model wherein the thresholds are the same across themes (Bond & Fox, 2020). In the context of quantitized themes, this model estimates several key parameters that help analyze how participants engage with themes that have been quantitized into ordinal categories (e.g., from "Strongly Disagree" to "Strongly Agree"). Specifically, it estimates threshold parameters, which are the points on the latent trait continuum where participants are likely to transition from one level of engagement or agreement to the next. In addition, it estimates theme difficulty, which is the overall difficulty or challenge associated with endorsing higher levels of engagement or agreement with each theme. Also, it estimates consistency across categories, which involves checking whether the rating scale is applied consistently across different themes, ensuring that the ordinal data are interpreted similarly across various contexts.

Multidimensional IRT Models

- Multidimensional 2PL IRT, Multidimensional 3PL IRT, Multidimensional 4PL IRT, and Multidimensional 5PL IRT, extend the 2PL IRT model, 3PL IRT model, 4PL IRT model, and 5PL IRT model, respectively, to multiple latent traits (abilities) (Reckase, 2009). For each theme,

- The Multidimensional 2PL IRT model estimates how difficult it is for participants to engage with the theme, considering multiple latent traits (e.g., different aspects of engagement). It also shows how well each theme differentiates between participants with varying levels of these traits.
- The Multidimensional 3PL IRT model, in addition to difficulty and discrimination, accounts for the possibility that participants might engage with a theme without fully understanding or being deeply engaged, by estimating a guessing parameter across multiple dimensions.
- The Multidimensional 4PL IRT model provides a comprehensive understanding of engagement by considering not just difficulty, discrimination, and guessing, but also the upper limit of how likely it is that participants will engage with a theme, even when they have high levels of the relevant traits.
- The Multidimensional 5PL IRT model provides the most nuanced analysis, considering not only difficulty, discrimination, guessing, and upper limits of engagement but also the asymmetry in how participants engage with themes. This model captures any skewed patterns in engagement, offering deep insights into how different participants interact with quantitized themes.

Hierarchical IRT Models

- **Higher-order IRT Models:** Higher-order IRT models are an extension of traditional IRT models that incorporate multiple levels of latent traits (Huang et al., 2013). They introduce a hierarchical structure whereby lower-order latent traits are influenced by one or more higher-order latent traits (Chen et al., 2006). In the context of quantitized themes, higher-order IRT models can help determine how a participant's general tendency to engage with themes (higher-order trait) influences their specific engagement with individual themes (lower-order traits). For example, a higher-order trait might represent general academic motivation, whereas lower-order traits could represent engagement with specific themes like "time management" or "peer collaboration." This model estimates the influence of broad, higher-order traits on specific, lower-order traits, helping to understand the hierarchical structure of engagement with quantitized themes, showing how specific engagements are connected and driven by broader tendencies.

- **Multilevel IRT Models:** Multilevel IRT models, also known as hierarchical IRT models, are designed to handle data that are nested within different levels, such as students within classrooms, employees within companies, or patients within hospitals (Bacci & Caviezel, 2011; Kamata & Cheong, 2007). These models extend traditional IRT by accounting for the fact that data might be influenced by factors at multiple levels (Reckase, 2009). In context of quantitized themes, multilevel IRT models can estimate how factors at different levels (e.g., school characteristics, teacher influence) impact engagement with themes. For example, a student's engagement with the theme of "collaborative learning" might be influenced by their individual traits (e.g., preference for group work) and also by the overall classroom environment. These models also allow for the comparison of engagement patterns across different groups (e.g., different schools or classes) while accounting for the nested structure of the data. This can provide insights into whether certain themes are more or less engaging depending on the broader context in which participants are situated.

Mixture IRT Models

- **Latent Class IRT Models:** Latent Class IRT models combine the principles of Latent Class Analysis (LCA) with IRT to identify distinct subgroups, or "latent classes," within the participant population based on their patterns of responses to quantitized themes (Rost, 1990). These models do not assume a continuous latent trait (like traditional IRT models) but, instead, estimate the probability that a participant belongs to a specific latent class (Collins & Lanza, 2009). Within each latent class, these models estimate item parameters such as difficulty and discrimination for the quantitized themes, but these parameters can vary among classes (Bacci et al., 2014). This allows for the identification of different response behaviors across the identified classes. In the context of quantitized themes, Latent Class IRT models can identify and describe distinct subgroups of participants who engage with themes in different ways. For example, one class might find certain themes easy with which to engage, whereas another class finds them more difficult. This helps in understanding how different groups within the population interact with the themes differently. By identifying these latent classes, researchers can tailor interventions or support strategies to the specific needs and characteristics of each class, based on how they engage with the themes.

- **Mixed-Effect IRT Models:** Mixed-Effect IRT models estimate both fixed effects (which apply to the entire population) and random effects (which vary across groups or individuals) (Rijmen et al., 2003). In the context of quantitized themes, these models allow for the estimation of how individual-level and group-level factors influence participants' engagement with themes. These models estimate variability in theme engagement at both the individual level and the group level (Kamata & Vaughn, 2011). For example, a model might estimate how much of the variation in engagement with a theme is due to individual differences and how much is due to differences between groups (e.g., different classes, schools, or demographic groups). Mixed-Effect IRT models are useful for understanding how context (e.g., group membership or environmental factors) influences participants' engagement with themes. For instance, the model might reveal that certain themes are engaged with differently depending on the classroom environment or socio-economic status. These models particularly are valuable in studies wherein data are nested or hierarchical, such as students within schools or employees within departments. They provide a nuanced analysis that accounts for both the individual and contextual influences on theme engagement.

Bayesian IRT Models

- **Bayesian Estimation in IRT:** Bayesian IRT Models involve the application of Bayesian statistical methods to IRT, offering a flexible framework for estimating various parameters associated with quantitized themes (Fox, 2010). In the context of quantitized themes, Bayesian IRT models estimate the same types of parameters as do traditional IRT models (e.g., difficulty, discrimination, guessing) but with several key differences and advantages due to the Bayesian approach. Bayesian IRT models estimate the difficulty of each quantitized theme, representing how challenging it is for participants to engage with that theme. These models also estimate how well each theme discriminates among participants with different levels of the underlying trait (e.g., engagement). Depending on the complexity of the Bayesian IRT model used (e.g., 2PL, 3PL, 4PL, or 5PL), additional parameters like guessing, upper asymptote, and asymmetry also can be estimated. Bayesian IRT models generate posterior distributions for all estimated parameters, including both theme parameters (like difficulty and discrimination) and participant engagement levels (latent traits). These distributions reflect the uncertainty about each parameter given the data and the prior information. For quantitized themes, this means that researchers can see not just the most likely values for difficulty or discrimination, but also the range of plausible values, given the data. This provides a more comprehensive picture of how confident we can be in the estimates.

  Bayesian IRT models estimate participants' latent traits (e.g., their overall engagement with the themes) using a probabilistic approach (Fox, 2010). These traits are estimated based on the observed responses to the themes, with the model accounting for both the data and any prior information about the traits (Fox & Glas, 2001). For quantitized themes, this means estimating each participant's engagement level, with the Bayesian approach providing a full posterior distribution for each participant's trait, reflecting the uncertainty around the estimate.

Bayesian IRT models allow for the integration of prior information or expert knowledge into the estimation process. This prior information can come from previous studies, expert judgments, or other sources. In the analysis of quantitized themes, this could mean incorporating previous knowledge about the difficulty of certain themes or the expected distribution of engagement levels among participants, thereby improving the accuracy of the estimates when the data are sparse or uncertain.

Bayesian methods provide tools for model comparison, such as Bayes factors or Deviance Information Criterion (DIC), to evaluate how well different models fit the data (Spiegelhalter et al., 2002). This allows researchers to compare different IRT models applied to quantitized themes (e.g., 2PL vs. 3PL) and to select the one that best represents the data, incorporating both the data and prior knowledge (Vehtari & Ojanen, 2012).

Bayesian IRT models particularly are advantageous in situations with small sample sizes, wherein traditional methods might be challenging (Levy & Mislevy, 2017). The use of prior information helps stabilize estimates when data are limited. The Bayesian approach offers flexibility in model specification, allowing researchers to incorporate complex structures and prior knowledge that might not be accommodated easily within traditional IRT models. By providing full posterior distributions, Bayesian IRT models give a more complete picture of uncertainty, which is particularly useful in understanding the robustness of the findings related to quantitized themes.

Diagnostic Classification Models (DCMs)

- **Cognitive Diagnostic Models (CDMs):** Cognitive Diagnostic Models (CDMs) are advanced psychometric models used primarily to diagnose specific skills or cognitive attributes that individuals possess or lack (cf. Leighton & Gierl, 2007). When applied to quantitized themes in mixed methods research, these models estimate how well participants engage with or understand specific themes based on underlying cognitive attributes or skills. These models estimate the presence or absence of specific cognitive attributes (skills or abilities) that are required to engage with or to understand quantitized themes. Each theme is associated with one or more attributes, and the model estimates whether a participant has these attributes. Based on the cognitive attributes, they estimate whether a participant is likely to master or to engage with a particular theme. The model provides a diagnostic profile for each participant, showing which attributes they possess and how these attributes contribute to their ability to engage with themes. CDMs can be used to identify specific strengths and weaknesses in participants' engagement with themes. For example, a theme might require several underlying skills, and the model will estimate which skills a participant has mastered and how these influence their engagement with the theme.

- **DINA Model (Deterministic Input, Noisy "And" gate):** The DINA model estimates whether a participant possesses all the necessary cognitive attributes required to engage with a quantitized theme (De La Torre, 2009). In this model, a participant must have mastered all relevant attributes to engage with the theme successfully; if even one attribute is missing, the model predicts failure in engagement. The model also estimates the probability that a participant will engage with a theme based on their mastery of the required attributes, taking into account potential "noise" or error in the process. The DINA model particularly is useful for themes that require a combination of skills or knowledge to be understood or engaged with fully. It assumes a conjunctive (AND) relationship, wherein all necessary attributes must be present for successful engagement with the theme. This model can highlight areas where participants are struggling because they lack one or more critical attributes.

- **DINO Model (Deterministic Input, Noisy "Or" gate):** The DINO model estimates whether a participant possesses at least one of the necessary cognitive attributes required to engage with a quantitized theme (Sessoms & Henson, 2018). In this model, a participant only needs to have mastered one relevant attribute to engage with the theme successfully; even partial mastery of the attributes can lead to engagement. Similar to the DINA model, the DINO model estimates the probability that a participant will engage with a theme based on their mastery of at least one of the required attributes, also considering noise or error. The DINO model is useful for themes wherein engagement can occur even if the participant only has partial knowledge or skills. It assumes a disjunctive (OR) relationship, wherein possessing any one of the relevant attributes can lead to successful engagement with the theme. This model can reveal situations wherein participants are able to engage with themes in different ways based on their diverse skill sets.

Conclusions Regarding Alternative IRT Models

In conclusion, the application of various IRT models—such as polytomous, multidimensional, hierarchical, mixture, Bayesian, and diagnostic classification models (DCMs)—offers a robust and multifaceted approach to analyzing quantitized themes. Polytomous IRT models capture the complexity of themes with multiple response levels, whereas multidimensional IRT models provide insights into the interplay among different underlying traits. Hierarchical IRT models add depth by assessing both general and specific latent traits, allowing for a layered understanding of theme interaction. Mixture IRT models uncover latent subgroups, revealing diverse patterns of engagement across the population. Bayesian IRT models enhance analysis with flexibility and the incorporation of prior knowledge, making them particularly useful in complex or uncertain data environments. Lastly, DCMs offer precise diagnostic information on the specific skills or attributes required for theme engagement, identifying areas of strength and challenge. Collectively, these IRT models provide a comprehensive toolkit that enables researchers to delve deeply into the nuances of participant interactions with quantitized themes, ultimately leading to more precise, insightful, and actionable conclusions.

Applying Rasch Analysis and Item Response Theory Models to Quantitized Qualitative Data: A Value-Added Approach to Inferential-Based Quantitizing

Differential Item Functioning in Rasch Analysis and IRT Analysis

Differential Item Functioning (DIF) within a Rasch analysis examines whether different groups of participants (e.g., by gender, age, or other demographics) respond differently to particular themes after controlling for overall levels of engagement (Chen & Revicki, 2023). More specifically, DIF analysis identifies whether certain themes are easier or more difficult for specific groups to endorse, indicating potential bias or differences in how themes are understood or engaged with by different groups (Chen & Revicki, 2023). By detecting and accounting for DIF, researchers can ensure that comparisons between groups are fair and not confounded by differences in how themes function across groups (Zumbo, 2007). This approach allows researchers to explore how different demographic groups engage with themes differently, adding an inferential layer to the interpretation of qualitative data.

DIF also can be explored within the broader IRT framework, specifically, the 2PL IRT, 3PL IRT, 4PL IRT, or 5PL IRT models (Lord, 1980). For example, DIF analysis in IRT models can reveal whether themes discriminate differently between groups or if the likelihood of contributing to a theme varies by group, even after accounting for participant ability. This can be particularly useful for understanding group-specific dynamics within the data. By identifying themes that function differently across groups, researchers can make more informed interpretations about the nature of the themes and their relevance to different subsets of participants. In this way, DIF analysis in IRT models provides a powerful tool for understanding and addressing potential biases in the qualitative data, ensuring that the insights derived from the data are robust and applicable across different groups.

Comparison of Differential Item Functioning Quantitizing of Themes and Inferential-Based Quantitizing

DIF of themes—which represent measurement-based quantitizing—and inferential-based quantitizing approaches, such as those using the GLM, are both used with the aim of understanding how different groups of participants engage with themes; but they do so in distinct ways. Specifically, DIF analysis focuses on whether specific themes function differently across demographic groups or other groups after accounting for each participant's general propensity or likelihood to engage with the themes identified in the qualitative data, which represents an estimate of each participant's overall level of engagement or contribution across all themes being analyzed. For instance, DIF can detect whether participants from a certain demographic group (e.g., based on age, gender, ethnicity) are less likely to contribute to that theme compared to participants from other groups, even after accounting for their general level of engagement—this would indicate potential bias or differential interpretation. Thus, DIF operates at the theme level, assessing each theme's fairness (i.e., under the assumption that each theme should be equally accessible and relevant to all participants, regardless of their demographic characteristics or any other group-defining characteristic) and ensuring that comparisons across groups are not confounded by these biases. It allows for the identification of subtle differences in how themes resonate with various groups, providing insight into the nuances of group-specific engagement with qualitative themes.

In contrast, inferential-based quantitizing approaches, which include GLM analyses (e.g., multiple regression, canonical correlation analysis) involve the treatment of themes as dependent variables and group-defining characteristic variables as independent variables. These GLM analyses examine the relationship between themes and group-defining factors at a more aggregate level, focusing on the statistical associations between the set of themes and the demographic predictors. Although this approach can highlight overall trends and relationships, it does not reveal the finer details of how *individual* themes function differently across groups. And even when inferential-based quantitizing approaches, such as chi-square analyses, are used to examine how individual themes function differently across demographic groups, there is still value-addedness to using DIF analysis over these such inferential-based methods for several reasons.

First, DIF controls for participants' overall level of engagement, meaning that it takes into account a participant's general level of engagement across all themes before assessing differences in how specific themes function for different groups. This is crucial because it isolates the effect of group membership (e.g., gender) on engagement with a specific theme, independent of the participant's general tendency to contribute to themes. Although chi-square analysis can reveal whether there is an association between a group-defining variable and a theme, it does not control for participants' overall level of engagement. This means that any observed association could be confounded by differences in general engagement levels between groups, potentially leading to misleading conclusions.

Second, DIF is designed specifically to detect subtle biases in how themes function across different groups. It can identify themes that appear *fair* (i.e., each theme being equally accessible and relevant to all participants) at an aggregate level but function differently for specific groups after accounting for overall level of engagement. For instance, a theme might be equally common across groups, but DIF could reveal that it is more challenging for one group to engage with meaningfully—that is, a specific theme is less accessible, less relevant, or more difficult to interpret and to engage with in a deep or significant way for one demographic group compared to others, even when their overall ability to engage with themes is similar. Univariate analyses such as chi-square tests are useful for identifying marked differences but might not detect more nuanced forms of bias. For example, a chi-square analysis might show no statistically significant association between gender and a theme, but DIF could reveal that, even when the gender groups have the same overall level of engagement or ability to contribute to themes, one gender (e.g., women) might still find a particular theme more difficult with which to connect, to interpret, or to contribute to compared to another gender (e.g., men). This suggests that, as revealed by the DIF analysis, the theme might be less relevant, less accessible, or more challenging for that gender—women in this example—independent of their general capacity to engage with themes overall. Such a difference indicates that the theme does not function equivalently across genders, potentially due to bias or differences in how the theme resonates with each group. For example, if men and women both have similar overall engagement levels, but men consistently contribute less to a theme about "work-life balance," the theme might be more challenging or less relevant for men despite their similar overall ability to engage with other themes. DIF also provides insights into the underlying functionality of themes within different group-defining (e.g., demographic) contexts, helping researchers understand not only whether a theme is associated with a particular group-defining factor, but also how and why it might be more or less accessible or relevant to different groups. As such, DIF provides a more granular analysis by focusing on each theme individually, allowing researchers to pinpoint specific themes that might be interpreted differently by various (demographic) groups. This level of detail often is lost in univariate analyses like chi-square analysis, which might overlook such nuances when analyzing themes individually.

Third, DIF can be applied within more complex IRT models (i.e., 2PL, 3PL, 4PL, 5PL) that consider multiple parameters. This use of DIF within the broader IRT framework allows researchers to explore complex interactions among participant engagement, theme difficulty, discrimination, guessing, and other parameters. This adds depth to the analysis by considering multiple facets of how themes engage different groups, which goes beyond the capabilities of traditional GLM approaches. This allows for a multidimensional understanding of how themes function across groups, which is beyond the scope of a basic chi-square analysis. Chi-square is a bivariate analysis, which means that it involves examining the relationship between two variables at a time. It does not account for the complexity of multiple contributing factors—such as discrimination or guessing in the context of IRT—that could affect how themes are interpreted by different groups.

Fourth, by ensuring that themes function equivalently across groups, DIF analysis contributes to measurement invariance. This means that the measurement of engagement with themes is consistent across different demographic groups, allowing for fair and valid comparisons. Chi-square analysis does not directly test for measurement invariance. It can indicate whether there is an association, but it does not confirm whether the theme is being interpreted and engaged with in the same way across groups.

Summary of Value-Added Nature of Differential Item Functioning Quantitizing of Themes Over Inferential-Based Quantitizing

In summary, although both DIF and inferential-based quantitizing provide valuable insights into the relationship between themes and group-defining (e.g., demographic) variables, DIF offers a more detailed, fair, and context-sensitive analysis. It allows researchers to detect and to adjust for biases at the theme level, ensuring that the qualitative data are interpreted robustly and equitably across different participant groups. This makes DIF a powerful tool for enhancing the trustworthiness of research findings in ways that both aggregate and univariate inferential-based quantitizing approaches might not fully achieve.

Alignment with Critical Dialectical Pluralism Principles through Enhanced Equity and Inclusivity. Critical dialectical pluralism 2.0 (CDP 2.0; Onwuegbuzie et al., 2024)) represents the latest evolution of the critical dialectical pluralism metaphilosophy developed by Onwuegbuzie and Frels (2013), integrating its foundational principles of **s**ocial justice, **i**nclusion, **d**iversity, **e**quity, and **s**ocial responsibility (SIDES) into every phase of the research process. These pillars collectively drive the mission of this multidimensional metaphilosphy to address systemic inequities and to produce research that fosters meaningful social change. In this context, DIF quantitizing of themes aligns with CDP 2.0 by providing a detailed, fair, and context-sensitive analysis. DIF ensures that biases in theme-level interpretations are identified and adjusted, enabling equitable representation of diverse participant groups. More specifically,

- **Social Justice**: DIF ensures that biases in qualitative data analysis are identified and mitigated, promoting fair and unbiased interpretations across participant groups. This resonates with CDP's commitment to dismantling systemic inequities and ensuring that research outcomes do not perpetuate existing power imbalances or marginalize groups.

- **Inclusion**: By incorporating context-sensitive analyses that accommodate diverse participant experiences, DIF aligns with CDP's emphasis on including historically marginalized voices. It supports a methodological framework wherein all participant perspectives are valued and integrated into the analysis.

- **Diversity**: DIF's capacity adaptively to analyze theme-level biases across demographic groups directly reflects CDP's pluralistic nature. This approach reflects the valuing of diverse epistemologies and perspectives, acknowledging the multifaceted nature of reality and the importance of addressing it in research.

- **Equity**: DIF's focus on ensuring equitable interpretation of data mirrors CDP's objective of flattening traditional power hierarchies in research. By adjusting for biases, DIF upholds the principle of fairness, ensuring that all groups are represented equitably.

- **Social Responsibility**: By enhancing the robustness and trustworthiness of research findings through detailed and context-sensitive analyses, DIF supports CDP's goal of making research findings actionable and socially beneficial. This reflects a commitment to producing research that has tangible, equitable impacts on the studied communities.

Psychometric Evaluation of Quantitized Themes

When applying IRT to the analysis of quantitized themes, three key tools—Test Information Function (TIF), Item Characteristic Curves (ICCs), and Item Fit Analysis—play an important role in evaluating and refining the measurement process. Each of these tools provides unique insights into different aspects of the thematic development, from how well themes discriminate among different levels of participant engagement with themes to how consistently they align with the underlying measurement model. By integrating these analyses, researchers can ensure that the quantitized themes accurately capture the intended constructs, ultimately enhancing the validity, reliability, and trustworthiness of the measurement-based quantitizing process.

Test Information Function (TIF)

- **Purpose:** In the context of themes, the TIF can be used to determine the reliability and informativeness of themes across different levels of participant engagement with themes, helping researchers to identify wherein themes provide the most or least useful data (De Ayala, 2013; Lord, 1980). The TIF provides insights into how much information the set of themes provides across different levels of participant engagement. It helps in understanding the reliability of the theme assessment across different engagement levels.

- **Application:** In the context of themes, TIF can reveal at which levels of participant engagement the themes are most informative and where they might lack reliability.

Item Characteristic Curves (ICCs)

- **Purpose:** In the context of themes, ICCs offer a visual representation of how the likelihood of engaging with a theme varies with participant overall engagement, providing detailed insights into the behavior of individual themes (Hambleton et al., 1991). ICCs plot the probability of engaging with a theme against participant overall engagement, showing how the likelihood of engaging with a theme changes as overall engagement increases (De Ayala, 2013; Lord, 1980).

- **Application:** For themes, ICCs can help visualize and understand how engagement with specific themes varies across different levels of participant engagement, providing a more detailed look at individual theme behavior.

Item Fit Analysis

- **Purpose:** Item Fit Analysis examines how well each theme conforms to the expected IRT model, ensuring that the themes reliably measure the intended latent traits (De Ayala, 2013;

Lord, 1980). Poorly fitting themes might suggest that the theme is not behaving consistently across different levels of participant overall engagement.

- **Application:** This analysis is critical in ensuring that each theme is valid and reliable, allowing, where applicable, researchers to identify and potentially to revise or to remove themes that do not conform well to the model.

Optimizing Theme Quantity and Sample Size for Reliable IRT Analysis of Quantitized Themes

In IRT analyses, in the context of quantitized themes, ensuring an adequate number of themes and an appropriate sample size is crucial for the reliability and validity of the results, particularly across the aforementioned models: Rasch, 1PL IRT, 2PL IRT, 3PL IRT, 4PL IRT, 5PL IRT, polytomous IRT models, multidimensional IRT models, hierarchical IRT models, mixture IRT models, Bayesian IRT models, and diagnostic classification models (DCMs). The number of themes influences the precision with which the latent traits are measured, especially in complex models or multidimensional IRT models, wherein multiple parameters (e.g., difficulty, discrimination, guessing) need to be estimated. Analyzing too few themes can lead to unstable parameter estimates, increased standard errors, and reduced power to detect differences or associations (De Ayala, 2013; Lord, 1980).

Similarly, the sample size plays a critical role in the accuracy of parameter estimation across different IRT models. Smaller sample sizes can lead to biased estimates, particularly in models with more complex parameter structures such as the 3PL IRT, 4PL IRT, 5PL IRT, and mixture IRT models (Lord, 1980). Adequate sample size also is critical in ensuring the generalizability of the findings, reducing the likelihood of overfitting, and improving the precision of parameter estimates (Karabatsos, 2015; König et al., 2020).

With these considerations in mind, in what follows, criteria will be provided for both the minimum number of themes and the minimum sample size required for each of the aforementioned IRT analyses, to ensure the robustness and validity of the results. These guidelines will help researchers design their studies with sufficient power and precision, avoiding the pitfalls associated with underestimating either factor. By adhering to these criteria, the accuracy of parameter estimation and the generalizability of findings can be improved significantly.

Minimum Number of Quantitized Themes and Participants for Reliable IRT Analysis

IRT Analysis on Less Than Four Quantitized Themes. Conducting IRT analyses on fewer than four items severely limits the model's ability accurately to estimate item parameters (e.g., difficulty, discrimination) and latent traits (e.g., ability levels). This is true for a wide range of IRT models, including the Rasch model, 1PL IRT, 2PL IRT, 3PL IRT, 4PL IRT, 5PL IRT, polytomous IRT models, multidimensional IRT models, hierarchical IRT models, mixture IRT models, Bayesian IRT models, and diagnostic classification models (DCMs). With fewer than four items, these models do not have sufficient data to capture the variability and nuances in the responses, leading to unstable and unreliable estimates. This issue is particularly acute in models like 2PL IRT, 3PL IRT, and beyond, wherein additional parameters, such as item discrimination and guessing, are involved (Galdin & Laurencelle, 2010). For these models, the data provided by fewer than four items often are insufficient to estimate the parameters accurately, which can result in misleading conclusions. Even in simpler models like the Rasch model or 1PL IRT, the lack of sufficient items can prevent the model from adequately distinguishing among individuals' levels of engagement, leading to poor model fit and less meaningful interpretations (Smith et al., 2008).

Even with a large sample size, conducting IRT analyses with fewer than four items remains unjustified. Although increasing the sample size can help stabilize parameter estimates to some extent, it cannot compensate for the inherent limitations imposed by the small number of items. The fundamental problem is that, with fewer than four items, the model lacks the necessary complexity and information to differentiate between the various dimensions of the latent trait being measured. This limitation applies across all IRT models, including polytomous and multidimensional IRT models, hierarchical models, and Bayesian models. For diagnostic classification models, which often are used to classify respondents into different latent classes based on their responses, an insufficient number of items can compromise severely the model's ability accurately to classify individuals, even if the sample size is large. Therefore, the quality and richness of the data, in terms of the number of items, are crucial for the success of these analyses.

These limitations regarding the analysis of fewer than four items also apply directly to the context of quantitized themes, which are themes that have been converted from qualitative data into quantitative variables. In measurement-based quantitizing, if fewer than four themes emerge from the qualitative analysis, the use of IRT models is not appropriate. Just as with traditional items, quantitized themes need to be numerous enough to provide the necessary data for stable and reliable parameter estimates. With fewer than four themes, the models lack the complexity required to capture the full range of variability and to provide meaningful insights into the underlying latent traits. This can lead to inaccurate or misleading interpretations, particularly in more complex models like the 2PL IRT, 3PL IRT, 4PL IRT, and 5PL IRT models, wherein themes play a crucial role in discriminating among different levels of the latent trait. Therefore, in order to ensure the validity and reliability of IRT analyses, it is essential that at least four quantitized themes are available because this allows the models to function properly and to yield interpretable results.

IRT Analysis on Four Quantitized Themes. When conducting an IRT analysis on four quantitized themes, for the simplest models—namely, Rasch or 1PL IRT—the utility of these analyses improves compared to scenarios with fewer themes. Although four themes are still on the lower end, these models can yield useful insights under certain conditions, particularly when the sample size is sufficiently large and the themes are well-targeted. Specifically, the Rasch model reasonably can be applied with four themes, particularly when the sample size is large enough (typically 500 or more participants) to ensure the stability of parameter estimates (Smith et al., 2008). With four themes, the model can provide meaningful measures of theme difficulty and person engagement, but the precision of these estimates might still be limited. The model's utility in this context lies in its simplicity and the capacity to provide a straightforward analysis of the relationship between themes and the latent trait being measured. The 1PL IRT model, which focuses on estimating theme difficulty, also can be applied with four themes. The utility of the 1PL IRT model with four themes is somewhat limited, but it can still be effective in contexts wherein the themes are well-targeted and the sample size is sufficient (500 or more participants). The model's utility increases if the items have a good spread of difficulty levels, which allows for a more accurate estimation of the latent trait (De Ayala, 2013). For more general IRT models, such as the 2PL IRT and 3PL IRT, conducting analyses with four themes is possible but remains somewhat limited (De Ayala, 2013). These models require sufficient data to estimate multiple parameters (e.g., difficulty, discrimination), and, with only four themes, the precision of these estimates might be compromised. However, when working with a large sample size (at least 500 participants), these models can still offer valuable insights into the relationships between themes and the underlying latent trait.

IRT Analysis on 5 to 9 Quantitized Themes. When considering the application of various IRT models to at least five quantitized themes, whether measured dichotomously or on a rating scale, the decision to conduct these analyses must be evaluated carefully based on the specific model, the characteristics of the data, and the intended outcomes of the analysis. The interaction between this number of themes and sample size is critical because insufficient items can result in poor parameter estimation, especially in more complex models that require adequate data points to estimate additional parameters accurately (Sébille et al., 2010). In what follows, each of the aforementioned IRT analyses is discussed with regard to their appropriateness for analyzing as five themes.

Rasch and 1PL IRT can be applied justifiably to as few as five themes, providing that a sufficient sample size is used—typically 500 or more participants. With five themes, these models can provide meaningful estimates of theme difficulty and participant level of engagement. However, the precision of these estimates still might be limited, and the model's utility might be constrained to contexts wherein simplicity and basic theme difficulty estimation are the primary goals (Smith et al., 2008). Indeed, Uyigue and Orheruata (2019) provided empirical evidence that “for an accurate item-difficulty parameter estimate in the 1PLM at least a test length of 10 and sample size of 1000 is required” (p. 72).

Polytomous IRT models, which assume a single population wherein all respondents are modeled with the same theme parameters and deal with themes with multiple response categories, can be appropriately applied to as few as five themes if these themes are measured on a rating scale with sufficient response categories. The effectiveness of these models with five themes depends on the nature of the data and the distribution of responses across categories. With an adequate sample size, these models can provide valuable insights, although the limited number of themes might constrain the depth of analysis (He & Wheadon, 2013; Mair & Hatzinger, 2007). However, for polytomous IRT models—which operate under the assumption that the population might be heterogeneous, with different latent subpopulations, each potentially having different theme parameters—at least 2,500 observations are required for accurate parameter and standard error estimates (Kutscher et al., 2019).

The 2PL IRT and 3PL IRT models, which introduce additional parameters like theme discrimination and guessing, can be applied to as few as five themes, but the utility of these models is limited. A larger sample size is critical (e.g., 1,000 or more participants) to stabilize the additional parameters. With five themes, these models still can offer insights, but the results should be interpreted cautiously because the limited number of themes can affect the reliability of the discrimination and guessing parameter estimates (Harwell & Janosky, 1991).

Bayesian IRT models can technically be applied to as few as five themes, particularly if the sample size is large (1,000 participants at a minimum; Waller & Feuerstahler, 2017) and informative priors are used. This recommendation is based on the need to ensure that the Bayesian estimation process, particularly when using informative priors, can compensate for the limited number of themes by leveraging the larger sample size. However, the utility of these models is limited, and the results should be interpreted with caution due to the limited number of themes (Karabatsos, 2015). In contrast, König et al. (2020) documented that, for optimized Bayesian hierarchical two-parameter logistic models, accurate item parameter estimates and trait scores are obtained even in sample sizes as small as 100.

For more complex IRT models—specifically, 4PL IRT and 5PL IRT—applying them to nine themes or less generally is unjustified, regardless of sample size. These models introduce even more parameters, such as upper and lower asymptotes, which require a substantial number of themes to be estimated accurately. With fewer than 10 themes, the estimates for these parameters are likely to be unstable and unreliable, making the analysis inappropriate (Cuhadar, 2022).

Similarly, multidimensional IRT models, which are designed to estimate multiple latent traits simultaneously, generally are inappropriate to apply to nine themes or less. These models require a larger number of themes to ensure stable parameter estimates. Analyzing fewer than 10 themes can lead to poor estimation of the latent traits and an inability adequately to capture the complexity of the multidimensional construct (i.e., the different dimensions of the latent traits). Applying these multidimensional IRT models to fewer than 10 themes can lead to poor model fit and unreliable parameter estimates, making the analysis unjustified regardless of sample size (Reckase, 2009).

Further, hierarchical IRT models, which involve multiple levels of analysis (e.g., items nested within groups), require a substantial number of themes to estimate parameters at each level accurately. Applying these models to as few as nine themes is inappropriate because it does not provide sufficient data to support the hierarchical structure of the model. Further, the model might not have sufficient data accurately to estimate the complex relationships among levels, leading to unreliable results (Huang et al., 2013).

In the same vein, mixture IRT models, which identify latent subpopulations within the data, generally are not justified with as few as nine themes. These models require a larger number of themes to distinguish between the different latent classes effectively. With only nine themes, the model might not be able to distinguish among different latent classes, leading to poor classification accuracy. Additionally, the ability accurately to estimate the parameters for each class is compromised, making the analysis inappropriate (Şen & Cohen, 2023).

Finally, diagnostic classification models, which classify respondents into different latent classes based on their responses, generally require more than nine themes to function effectively. With only nine themes, the model's ability accurately to classify respondents into the correct latent classes significantly is compromised, making the analysis inappropriate. This limitation arises because a smaller number of themes can lead to insufficient data for the reliable estimation of theme parameters and latent class membership, ultimately resulting in poor classification accuracy and unstable parameter estimates (Seo & Kim, 2021).

In summary, some IRT models—specifically, Rasch, 1PL IRT, and Polytomous IRT models (if the themes are measured on a rating scale with sufficient response categories)—can be applied to as few as five themes with a sufficient sample size. However, many other models, particularly those with more complex parameter structures, are inappropriate for use with nine themes or less. For models such as the 4PL IRT, 5PL IRT, multidimensional IRT, hierarchical IRT, mixture IRT, and DCMs, applying them to as few as nine themes is unjustified, regardless of the sample size, due to the significant limitations in parameter estimation and model reliability.

IRT Analysis on 10 to 14 Quantitized Themes. In addition to Rasch, 1PL IRT, 2PL IRT, 3PL IRT, and potentially Bayesian IRT models being appropriate for analyzing as few as five themes, polytomous IRT models can sometimes function with as few as 10 themes, particularly if each theme has multiple response categories. However, the reliability of parameter estimates and the model's ability to capture the full range of latent traits might still be limited with only 10 themes (Davier & Yamamoto, 2003). Although polytomous IRT models can be applied to datasets with as few as 10 themes, this approach generally is more appropriate in scenarios wherein the themes are well-distributed across the response categories, and the sample size is sufficiently large (i.e., at least 250-300 participants) to compensate for the reduced number of themes, thereby minimizing potential biases and ensuring more stable estimates (Svetina Valdivia & Dai, 2024).

IRT Analysis on 15 to 19 Quantitized Themes. Multidimensional IRT models estimate multiple latent traits simultaneously. With 15 themes, these models might function adequately, especially if the data are well-structured and the model is not overly complex. However, caution is needed as the number of dimensions increases, which might require more themes for stable estimates (Huo et al., 2015).

IRT Analysis on 20 to 29 Quantitized Themes. For the 3PL IRT model, which includes item difficulty, discrimination, and guessing parameters, using 20 themes might be sufficient if the themes are well-distributed and the sample size is adequate. However, it is important to ensure that the themes have a broad range of difficulty levels to leverage fully the model's capacity for differentiating among respondents (Dodeen, 2004). Additionally, careful attention should be given to the quality of the data to minimize the risk of overfitting, particularly with the guessing parameter (Partchev, 2009). In fact, via a simulation study, Djidu et al. (2023) proposed a minimum of 25 items for accurate estimation using the 3PL IRT model.

IRT Analysis on 30-34 Quantitized Themes. For the 4PL IRT and 5PL IRT models, 30 themes can be adequate if the sample size is large enough to support the estimation of additional parameters like upper and lower asymptotes. Specifically, with respect to the 4PL IRT model, a minimum sample size of approximately 4,000 is recommended for accurate item parameter recovery, including the estimation of the upper-asymptote (slipping) parameter (Cuhadar, 2022). This size ensures that the complex model parameters are estimated with sufficient precision (Cuhadar, 2022). Based on this recommendation, the 5PL IRT model, which adds even more complexity by including both lower and upper asymptote parameters, likely requires more than 4,000 to ensure accurate parameter estimation. Such a sample size would help in managing the additional complexity introduced by the fifth parameter and ensures that the model can be estimated reliably.

IRT Analysis on 35-39 Quantitized Themes. Mixture IRT models, which account for latent classes within the data, can be effective with 35 themes, particularly when the model is used to identify a reasonable number of latent classes and the sample size is sufficient to support the complexity of the analysis. Specifically, at least 2,000 study participants are needed to justify the analysis of 35 themes (Şen & Cohen, 2023). Additionally, larger sample sizes, such as 2,500 to 5,000 participants, might enhance further the accuracy and stability of parameter estimates, particularly as the number of latent classes or model complexity increases (Şen & Cohen, 2023).

IRT Analysis on 40-44 Quantitized Themes. Hierarchical IRT models involve multiple levels of data analysis, and, with 40 themes, these models can function effectively, especially in scenarios wherein the hierarchy is not overly complex. To ensure accurate parameter estimation and the stability of the model, a minimum sample size of approximately 3,000 to 5,000 participants likely is needed (cf. Huo et al., 2015). This sample size allows for more reliable parameter estimation across different levels of the hierarchy and can support more nuanced analyses of the data, enabling the model to capture subtle relationships and variations within the dataset (Huo et al., 2015).

IRT Analysis on 45 or More Quantitized Themes. As noted previously, DCMs are statistical models designed to classify examinees based on their mastery of a set of latent traits or skills (Ravand et al., 2020; Sessoms & Henson, 2018). With 45 themes, DCMs effectively can classify respondents, particularly when the model is designed for assessment purposes (cf. Henson, 2009). The number of themes should be sufficient to ensure reliable and valid classifications, especially in well-structured datasets wherein each theme contributes meaningfully to the diagnostic process (Sessoms & Henson, 2018). Specifically, with 45 themes, a minimum sample size of approximately 2,500 to 4,000 participants likely is needed. This range allows for accurate classification of respondents and reliable parameter estimation, ensuring that the model's complexity adequately is supported (Sessoms & Henson, 2018).

Summary and Conclusions Regarding Optimizing Theme Quantity and Sample Size for Reliable IRT Analysis of Quantitized Themes

The DCM represents the final IRT analysis in the set of aforementioned models, each of which requires careful consideration of theme quantity and sample size for reliable results. In the context of IRT analyses applied to quantitized themes, both the number of themes and the sample size play crucial roles in ensuring the validity and reliability of the results. The number of themes affects the precision with which latent traits are measured, particularly in more complex models like multidimensional IRT and mixture IRT models, wherein multiple parameters—such as difficulty, discrimination, and guessing—must be estimated accurately. A limited number of themes can lead to unstable parameter estimates, higher standard errors, and reduced power to detect differences or associations, making the analysis less effective (Paek & Cai, 2014).

Similarly, the adequacy of the sample size is paramount in achieving accurate parameter estimation across different IRT models. Smaller sample sizes can introduce bias, particularly in models with complex parameter structures such as the 3PL, 4PL, 5PL, and mixture IRT models. Larger sample sizes help mitigate these issues, ensuring generalizability, reducing the risk of overfitting, and improving the precision of the estimates (Şen & Cohen, 2023). In order to support robust and valid conclusions, it is essential to adhere to guidelines that dictate the minimum number of themes and the minimum sample size required for each specific IRT model.

In summary, the appropriateness of IRT models for analyzing quantitized themes depends heavily on both the number of themes and the sample size. As can be seen from Table 4—which provides a summary of the minimum number of themes and minimum number of participants for each IRT model—although simpler models like Rasch and 1PL IRT can be applied effectively to as few as five themes with a sufficient sample size, more complex models, such as the 4PL IRT, 5PL IRT, and multidimensional IRT models, generally require a larger number of themes and participants to ensure accurate and stable parameter estimation. Specifically, for DCMs, which are designed to classify examinees based on their mastery of latent traits, the analysis becomes robust with at least 45 themes and a sample size ranging from 2,500 to 4,000 participants, ensuring the validity and reliability of the classifications (Sessoms & Henson, 2018). These criteria help safeguard against the potential pitfalls of insufficient data, allowing researchers to conduct analyses that are both meaningful and reliable.

Table 4 clearly demonstrates that when three or fewer themes are extracted from a qualitative analysis, undertaking any form of IRT analysis is not warranted due to the insufficient data to support stable and reliable parameter estimation. Additionally, the table highlights that if a study has fewer than 250 participants, IRT analysis generally is inappropriate because the sample size is too small to yield meaningful results. However, when at least four themes are extracted and the sample size reaches a minimum of 500 participants, several forms of IRT analysis—particularly Rasch, 1PL IRT, 2PL IRT, and 3PL IRT—become justified, providing valuable insights into the data.

Moreover, for studies with at least 250 participants and between 10 and 34 themes, polytomous IRT can be conducted effectively, offering a robust analysis of data with multiple theme categories. Importantly, Table 4 underscores that, as both the number of themes and the number of participants increase beyond these minimum guidelines, a broader range of IRT analyses become available, allowing for more nuanced and comprehensive evaluations of the extracted qualitative themes. This relationship emphasizes the importance of ensuring sufficient themes and participants to maximize the potential of IRT analyses.

In conclusion, the emphasis on adequate theme quantity and sample size across IRT models highlights the necessity for careful study design involving quantitized themes. By adhering to these guidelines, researchers can enhance the accuracy of their analyses, ensuring that the results are robust, interpretable, and applicable in real-world settings. In so doing, they contribute to the ongoing advancement and refinement of measurement theory and the effective application of IRT models (Huo et al., 2015; Şen & Cohen, 2023) as part of the qualitative thematic process within a mixed methods research framework (Onwuegbuzie, 2024). This careful consideration of theme quantity and sample size is crucial not only for the precision of parameter estimates, but also for the generalizability and robustness of the findings, ensuring that the analyses yield meaningful and reliable insights into the underlying latent traits being measured.

Closing Thoughts

In the present article, I have provided a renewed call for expanding the practice of quantitizing in mixed methods research. First, I described Onwuegbuzie's (2024) categorization of the process of quantitizing into the following four distinct types: descriptive-based quantitizing, inferential-based quantitizing, exploratory-based quantitizing, and measurement-based quantitizing. Descriptive-based quantitizing involves converting qualitative data into quantitative measures, such as frequencies or percentages, to summarize the data. Inferential-based quantitizing extends this by using statistical methods to make inferences about populations based on quantitized data. Measurement-based quantitizing focuses on using quantitized data to assess and to measure latent traits, often through the application of psychometric models. Finally, exploratory-based quantitizing is used to explore patterns or relationships within quantitized data, often as a precursor to further quantitative analysis.

Among these four quantitizing processes, descriptive-based quantitizing is the most prevalent because it is often the first step in integrating qualitative and quantitative data in mixed methods research. In contrast, measurement-based quantitizing is the least commonly used. Therefore, the remainder of the manuscript has focused on expanding the concept of measurement-based quantitizing by outlining how modern test theory approaches—specifically, Rasch analysis and item response theory (IRT) models, can be applied to quantitized themes or even finer units such as categories, codes, or sub-codes.

Table 4

Minimum Number of Quantitized Themes and Participants for Reliable IRT Analysis

	Number of Themes						
IRT Model	0-3	4	5-9	10-14	15-19	20-29	30-34
Rasch Model	N/J	500+	500+	500+	500+	500+	500+
1PL IRT	N/J	500+	500+	500+	500+	500+	500+
2PL IRT	N/J	500+	1,000+	1,000+	1,000+	1,000+	1,000+
3PL IRT	N/J	500+	1,000+	1,000+	1,000+	1,000+	1,000+
4PL IRT	N/J	N/J	N/J	N/J	N/J	4,000+	4,000+
5PL IRT	N/J	N/J	N/J	N/J	N/J	4,000+	4,000+
Polytomous IRT	N/J	N/J	2,500+	250-300+	250-300+	250-300+	250-300+
Multidimensional IRT	N/J	N/J	1,000+	N/J	1,000+	N/J	N/J
Hierarchical IRT	N/J	N/J	N/J	N/J	N/J	N/J	N/J
Mixture IRT	N/J	N/J	N/J	N/J	N/J	N/J	N/J
Bayesian IRT	N/J	N/J	1,000+	1,000+	1,000+	1,000+	1,000+
Diagnostic Classification Models	N/J	N/J	N/J	N/J	N/J	N/J	N/J

Note: N/J = Not justified

In the article, I outline how applying Rasch analysis and foundational IRT models (e.g., 1PL IRT, 2PL IRT, 3PL IRT) to quantitized qualitative data offers a value-added approach to descriptive-based quantitizing. These models allow researchers to assess parameters (e.g., the difficulty and discrimination of themes) that provide deeper insights than do simple frequency counts. Moreover, other IRT models—in particular, 4PL IRT, 5PL IRT, polytomous IRT, multidimensional IRT, hierarchical IRT, mixture IRT, Bayesian IRT models, and Diagnostic Classification Models—offer even more sophisticated tools for analyzing quantitized data. Each model is discussed briefly in the context of its application to quantitized themes, showing how these advanced models can refine our understanding of the data by evaluating theme characteristics and their relationship to latent traits.

Furthermore, I discuss how Rasch analysis and IRT models contribute to inferential-based quantitizing by enabling differential item functioning (DIF) analysis. DIF analysis, which is consistent with the tenets of the CDP multidimensional metaphilosophy, identifies whether themes perform differently across subgroups, thereby adding an inferential dimension to quantitized qualitative data. Three key tools—Test Information Function (TIF), Item Characteristic Curves (ICCs), and Item Fit Analysis—are highlighted as being crucial for evaluating and refining the measurement process in the application of IRT models to quantitized themes. I emphasize the importance of optimizing the number of themes and the sample size for reliable IRT analysis by providing detailed guidelines on the minimum number of quantitized themes and participants required to conduct meaningful IRT analyses. For example, if fewer than four qualitative themes are extracted and quantitized or the sample size is below 250 participants, IRT analysis generally is not justified. However, when there are at least four themes and 500 participants, IRT models such as Rasch, 1PL IRT, 2PL IRT, and 3PL IRT become viable. Moreover, Polytomous IRT models can be employed with at least 250 participants and 10 to 34 themes, expanding the range of available analyses as both theme quantity and sample size increase.

I conclude this article by reiterating the potential of IRT models to enhance mixed methods research through measurement-based quantitizing. By applying these models to quantitized themes, researchers can achieve a more nuanced understanding of qualitative data, moving beyond mere descriptive statistics to more sophisticated analyses that assess the quality and performance of themes. Also, I highlight the role of DIF analysis in uncovering subgroup differences, further demonstrating the value of IRT models in inferential-based quantitizing.

I underscore that to leverage fully the benefits of IRT models, researchers must carefully consider the number of themes and sample size in their studies. In particular, a minimum of four themes and 500 participants are necessary to conduct reliable IRT analyses, with more complex models requiring larger datasets. This emphasis on adequate theme quantity and sample size is crucial for ensuring that IRT analyses yield robust and meaningful insights.

Overall, measurement-based quantitizing is a critical bridge between qualitative explorations of complex constructs and their rigorous, systematic quantification, enabling deeper insights and broader applications in scientific research. This methodological approach not only enriches the precision and interpretability of qualitative data, but also extends the reach of these findings into domains traditionally dominated by quantitative analysis. Furthermore, it fosters a more integrative and holistic understanding of research phenomena, allowing for the synthesis of qualitative depth with quantitative rigor in ways that enhance the validity and generalizability of research outcomes.

Unfortunately, I was unable to identify a single study wherein IRT models (e.g., Rasch model, 1PL IRT, 2PL IRT, 3PL IRT, 4PL IRT, 5PL IRT, polytomous IRT models, multidimensional IRT models, hierarchical IRT models, mixture IRT models, Bayesian IRT models, DCMs) were conducted *directly* on quantitized themes. Such measurement-based quantitizing would represent full(er) integration of qualitative and quantitative research approaches—what Onwuegbuzie (2017) coined as 1 + 1 = 1 integration (see also Hitchcock & Onwuegbuzie, 2022; Natesan et al., 2019; Newman et al., 2015; Onwuegbuzie, 2023; Onwuegbuzie & Hitchcock, 2019, 2022; Onwuegbuzie et al., 2018)—especially when conducting analyses like DIF. Therefore, I call for broader adoption of measurement-based quantitizing in mixed methods research, arguing that the integration of modern test theory approaches, like Rasch analysis and IRT models, can enrich significantly the analytical process. This integration not only allows for more precise measurement of qualitative constructs, but also facilitates the identification of subtle patterns and relationships that might be overlooked with traditional qualitative or quantitative methods alone. By expanding the use of these models, researchers can gain deeper insights into quantitized data, ultimately enhancing the rigor and impact of their studies.

References

Abt, K. (1987). Descriptive data analysis: A concept between confirmatory and exploratory data analysis. *Methods of Information in Medicine, 26*(2), 77-88. https://doi.org/10.1055/s-0038-1635488

Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms?. *Medical care*, *42*(1), I-7. https://doi.org/10.1097/01.mlr.0000103528.48582.7c

Bacci, S., Bartolucci, F., & Gnaldi, M. (2014). A class of multidimensional latent class IRT models for ordinal polytomous item responses. *Communications in Statistics-Theory and Methods*, *43*(4), 787-800. https://doi.org/10.1080/03610926.2013.827718

Bacci, S., & Caviezel, V. (2011). Multilevel IRT models for the university teaching evaluation. *Journal of Applied Statistics, 38*(12), 2775-2791. https://doi.org/10.1080/02664763.2011.570316

Barton, M. A., & Lord, F. M. (1981). An upper asymptote model for the three-parameter logistic item-response curves. *ETS Research Report Series, 1981*(1), i-8. https://doi.org/10.1002/j.2333-8504.1981.tb01239.x

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Bombatkar, A., & Parvat, T. (2015). Improvements in clustering using affinity propagation: A review. *Journal of Multidisciplinary Engineering Science and Technology*, *2*(6), 3159-0040. https://doi.org/10.1016/j.eswa.2014.09.054

Bond, T. G., & Fox, C. M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Psychology Press.

Bouguettaya, A., Yu, Q., Liu, X., Zhou, X., & Song, A. (2015). Efficient agglomerative hierarchical clustering. *Expert Systems with Applications*, *42*(5), 2785-2797. https://doi.org/10.1016/j.eswa.2014.09.054

Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational evaluation and policy analysis, 15*(2), 195-207. https://doi.org/10.3102/01623737015002195

Celebi, M. E., Kingravi, H. A., & Vela, P. A. (2013). A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert systems with applications*, *40*(1), 200-210. https://doi.org/10.1016/j.eswa.2012.07.021

Chen, W. H., & Revicki, D. (2023). Differential item functioning (DIF). In F. Maggino (Ed.) *Encyclopedia of quality of life and well-being research*. Springer

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bi-factor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189-225. https://doi.org/10.1207/s15327906mbr4102_5

Collins, L. M., & Lanza, S. T. (2009). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (Vol. 718). John Wiley & Sons.

Cottrell, M., Olteanu, M., Rossi, F., & Villa-Vialaneix, N. N. (2018). Self-organizing maps, theory and applications. *Revista de Investigacion Operacional*, *39*(1), 1-22.

Cuhadar, I. (2022). Sample size requirements for parameter recovery in the 4-parameter logistic model. *Measurement: Interdisciplinary Research and Perspectives*, *20*(2), 57-72. https://doi.org/10.1080/15366367.2021.1934805

David, S. L., Hitchcock, J. H., Ragan, B., Brooks, G., & Starkey, C. (2018). Mixing interviews and Rasch modeling: Demonstrating a procedure used to develop an instrument that measures trust. *Journal of Mixed Methods Research, 12*(1), 75-94. https://doi.org/10.1177/1558689815624586

Davier, M., & Yamamoto, K. (2003). Partially observed mixtures of IRT models: An extension of the generalized partial-credit model. *Applied Psychological Measurement, 28*(6), 389-406. https://doi.org/10.1177/0146621604268734

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford.

De La Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of educational and behavioral statistics*, *34*(1), 115-130. https://doi.org/10.3102/1076998607309474

Djidu, H., Retnawati, H., & Haryanto, H. (2023). Ensuring parameter estimation accuracy in 3PL IRT modeling: The role of test length and sample size. *JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia)*, *12*(2), 177-190. https://doi.org/10.15408/jp3i.v12i2.34130

Dodeen, H. (2004). The relationship between item parameters and item fit. *Journal of Educational Measurement, 41*(3), 261-270. https://doi.org/10.1111/J.1745-3984.2004.TB01165.X

Fischer, G. H. (1995). Derivations of the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 15-38). Springer.

Fox, J. P. (2010). *Bayesian item response modeling: Theory and applications*. Springer.

Fox, J. P., & Glas, C. A. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*, 271-288. https://doi.org/10.1007/BF02294839

Galdin, M., & Laurencelle, L. (2010). Assessing parameter invariance in item response theory's logistic two item parameter model: A Monte Carlo investigation. *Tutorials in Quantitative Methods for Psychology, 6*(2), 39-51. https://doi.org/10.20982/TQMP.06.2.P039

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems, 12*, 436-445. https://doi.org/10.1525/sp.1965.12.4.03a00070

Gottschalk, P. G., & Dunn, J. R. (2005). The five-parameter logistic: A characterization and comparison with the four-parameter logistic. *Analytical biochemistry*, *343*(1), 54-65. https://doi.org/10.1016/j.ab.2005.04.035

Hahn, L. W., Ritchie, M. D., & Moore, J. H. (2003). Multifactor dimensionality reduction software for detecting gene–gene and gene–environment interactions. *Bioinformatics*, *19*(3), 376-382. https://doi.org/10.1093/bioinformatics/btf869

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Harwell, M. R., & Janosky, J. E. (1991). An empirical study of the effects of small datasets and varying prior variances on item parameter estimation in BILOG. *Applied Psychological Measurement*, *15*(3), 279-291. https://doi.org/10.1177/014662169101500308

Hayat, B., Putra, M. D. K., & Suryadi, B. (2020). Comparing item parameter estimates and fit statistics of the Rasch model from three different traditions. *Jurnal Penelitian dan Evaluasi Pendidikan*, *24*(1), 39-50. https://doi.org/10.21831/pep.v24i1.29871

He, Q., & Wheadon, C. (2013). The effect of sample size on item parameter estimation for the partial credit model. *International Journal of Quantitative Research in Education, 1*(3), 297-315. https://doi.org/10.1504/IJQRE.2013.057692

Henson, R. A. (2009). Diagnostic classification models: Thoughts and future directions. *Measurement: Interdisciplinary Research and Perspectives, 7*(1), 34-36. https://doi.org/10.1080/15366360802715395

Hitchcock, J. H., & Onwuegbuzie, A. J. (2022). The Routledge handbook for advancing integration in mixed methods research: An introduction. In J. H. Hitchcock & A. J. Onwuegbuzie (Eds.), *Routledge handbook for advancing integration in mixed methods research* (pp. 3-27). Routledge.

Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2013). Multidimensional scaling. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(1), 93-103. https://doi.org/10.1002/wcs.1203

Huang, H. Y., Wang, W. C., Chen, P. H., & Su, C. M. (2013). Higher-order item response models for hierarchical latent traits. *Applied Psychological Measurement*, *37*(8), 619-637. https://doi.org/10.1177/0146621613488819

Huo, Y., de la Torre, J., Mun, E. Y., Kim, S. Y., Ray, A. E., Jiao, Y., & White, H. R. (2015). A hierarchical multi-unidimensional IRT approach for analyzing sparse, multi-group data for integrative data analysis. *Psychometrika, 80*, 834-855. https://doi.org/10.1007/s11336-014-9420-2

Kalkan, Ö. K. (2022). The comparison of estimation methods for the four-parameter logistic item response theory model. *Measurement: Interdisciplinary Research and Perspectives*, *20*(2), 73-90. https://doi.org/10.1080/15366367.2021.1897398

Kamata, A., & Cheong, Y. F. (2007). Multilevel IRT models. In S. Sinharay & B. W. Wollack (Eds.), *Handbook of statistics: Vol. 26. Psychometrics* (pp. 543-567). Elsevier.

Kamata, A., & Vaughn, B. K. (2011). Multilevel IRT modeling. In A. Kamata & B. K. Vaughn (Eds.) *Handbook of advanced multilevel analysis* (pp. 41-57). Routledge.

Karabatsos, G. (2015). A Bayesian nonparametric IRT model. *arXiv: Methodology*.

König, C., Spoden, C., & Frey, A. (2020). An optimized Bayesian hierarchical two-parameter logistic model for small-sample item calibration. *Applied Psychological Measurement, 44*(4), 311-326. https://doi.org/10.1177/0146621619893786

Koskey, K. L. K., Sondergeld, T. A., Stewart, V. C., & Pugh, K. J. (2018). Applying the mixed methods instrument development and construct validation process: The Transformative Experience Questionnaire. *Journal of Mixed Methods Research, 12*(1), 95-122. https://doi.org/10.1177/1558689816633310

Koskey, K. L. K., & Stewart, V. C. (2014). A concurrent mixed methods approach to examining the quantitative and qualitative meaningfulness of absolute magnitude estimation scales in survey research. *Journal of Mixed Methods Research, 8*(2), 180-202. https://doi.org/10.1177/1558689813496905

Kramer, J. M. (2011). Using mixed methods to establish the social validity of a self-report assessment: An illustration using the Child Occupational Self-Assessment (COSA). *Journal of Mixed Methods Research, 5*(1), 52-76. https://doi.org/10.1177/1558689810386376

Kutscher, T., Eid, M., & Crayen, C. (2019). Sample size requirements for applying mixed polytomous item response models: Results of a Monte Carlo simulation study. *Frontiers in Psychology*, *10*, 2494. https://doi.org/10.3389/fpsyg.2019.02494

Leach Sankofa, N. (2022). Transformativist measurement development methodology: A mixed methods approach to scale construction. *Journal of Mixed Methods Research, 16*(3), 307-327. https://doi.org/10.1177/15586898211033698

Lee, D. D., & Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In T. Leen, T. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems, 13*, 556-562. MIT Press.

Leighton, J., & Gierl, M. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.

Le Roux, B., & Rouanet, H. (2010). *Multiple correspondence analysis* (Vol. 163). Sage.

Levy, R., & Mislevy, R. J. (2017). *Bayesian psychometric modeling*. Chapman and Hall/CRC.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*(1), 21-39. https://doi.org/10.1037/1082-989X.10.1.21

Madeira, S. C., & Oliveira, A. L. (2004). Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM transactions on computational biology and bioinformatics*, *1*(1), 24-45. https://doi.org/10.1109/TCBB.2004.2

Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, *20*, 1-20. https://doi.org/10.18637/JSS.V020.I09

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174. https://doi.org/10.1007/BF02296272

McClure, D. R., Ojo, E. O., Schaefer, M B., Bell, D., Abrams, S. S., & Onwuegbuzie, A. J. (2021). Online learning challenges experienced by university students in the New York City area during the COVID-19 pandemic: A mixed methods study. *International Journal of Multiple Research Approaches, 13*(2), 150-167. https://doi.org/10.29034/ijmra.v13n2editorial4

McDonald, R. P. (2013). Modern test theory. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. 1, pp. 118-143). Oxford University Press.

Michailidis, G. (2007). Correspondence analysis. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 191-194). Sage.

Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage.

Morell, L., & Tan, R. J. B. (2009). Validating for use and interpretation: A mixed methods contribution illustrated. *Journal of Mixed Methods Research, 3*(3), 242-264. https://doi.org/10.1177/1558689809335079

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159-176. https://doi.org/10.1177/014662169201600206

Murtagh, F. (1983). A survey of recent advances in hierarchical clustering algorithms. *The computer journal*, *26*(4), 354-359. https://doi.org/10.1093/comjnl/26.4.354

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of classification*, *31*, 274-295. https://doi.org/10.1007/s00357-014-9161-z

Natesan, P., Onwuegbuzie, A. J., Hitchcock, J., & Newman, I. (2019). Fully Integrated Bayesian thinking: A mixed methods approach to the 1 + 1 = 1 formula. *AERA Division D Newsletter*, 10-12. http://www.aera.net/Portals/38/docs/DivD/DNews_current/DivDNewsletter_Spring19.pdf

Newman, I., Onwuegbuzie, A. J., & Hitchcock, J. H. (2015). Using the general linear model to facilitate the full integration of qualitative and quantitative analysis: The potential to improve prediction and theory building and testing. *General Linear Model Journal, 41*(1), 12-28. http://www.glmj.org/archives/articles/Newman_v41n1.pdf

Onwuegbuzie, A. J. (2003). Effect sizes in qualitative research: A prolegomenon. *Quality & Quantity: International Journal of Methodology, 37*, 393-409. https://doi.org/10.1023/A:1027379223537

Onwuegbuzie, A. J. (2017, March). *Mixed methods is dead! Long live mixed methods!* Invited keynote address presented at the Mixed Methods International Research Association Caribbean Conference at Montego Bay, Jamaica.

Onwuegbuzie, A. J. (2021). Beyond identifying emergent themes in mixed methods research studies: The role of economic indices: The Thematic Herfindahl-Hirschman Index and the Thematic Concentration Ratio. *International Journal of Multiple Research Approaches, 13*(2), 137-149. https://doi.org/10.29034/ijmra.v13n2editorial3

Onwuegbuzie, A. J. (2022). Towards full(er) integration in mixed methods research: The role of canonical correlation analysis for integrating quantitative and qualitative data. *Publicaciones, 52*(2), 11-34. https://doi.org/10.30827/publicaciones.v52i2.27664

Onwuegbuzie, A. J. (2023). The 1 + 1 = 1 and 1 + 1 = 3 Integration formulas in mixed methods research:

A poem promoting peaceful and productive co-existence. *Journal of Mixed Method Studies, 8*, 17-22. https://doi.org/10.14689/jomes.2022.7.X

Onwuegbuzie, A. J. (2024). On quantitizing revisited. *Frontiers in Psychology, 15*, 1421525. https://doi.org/10.3389/fpsyg.2024.1421525

Onwuegbuzie, A. J., Abrams, S. S., & Forzani, E. (2024). Critical dialectical pluralism 2.0: A multidimensional metaphilosophy addressing social justice, inclusion, diversity, equity, and social responsibility. *International Journal of Multiple Research Approaches, 16*(3).

Onwuegbuzie, A. J., Bustamante, R. M., & Nelson, J. A. (2010). Mixed research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research*, *4*, 56-78. https://doi.org/10.1177/1558689809355805

Onwuegbuzie, A. J., & Combs, J. P. (2010). Emergent data analysis techniques in mixed methods research: A synthesis. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (2nd ed., pp. 397-430). Sage.

Onwuegbuzie, A. J., & Frels, R. K. (2013). Introduction: Toward a new research philosophy for addressing social justice issues: Critical dialectical pluralism 1.0. *International Journal of Multiple Research Approaches, 7*(1), 9-26. https://doi.org/10.5172/mra.2013.7.1.9

Onwuegbuzie, A. J., Frels, R. K., Leech, N. L., & Collins, K. M. T. (2011). A mixed research study of pedagogical approaches and student learning in doctoral-level mixed research courses. *International Journal of Multiple Research Approaches, 5*, 169-199. https://doi.org/10.5172/mra.2011.5.2.169

Onwuegbuzie, A. J., & Hitchcock, J. H. (2019). Toward a fully integrated approach to mixed methods research via the 1 + 1 = 1 integration approach: Mixed Research 2.0. *International Journal of Multiple Research Approaches, 11*(1), 7-28. https://doi.org/10.29034/ijmra.v11n1editorial1

Onwuegbuzie, A. J., & Hitchcock, J. H. (2022). Towards a comprehensive meta-framework for full integration in mixed methods research. In J. H. Hitchcock & A. J. Onwuegbuzie (Eds.), *Routledge handbook for advancing integration in mixed methods research* (pp. 565-606). Routledge.

Onwuegbuzie, A. J., Hitchcock, J. H., Natesan, P., & Newman, I. (2018). Using fully integrated Bayesian thinking to address the 1 + 1 = 1 integration challenge. *International Journal of Multiple Research Approaches*, *10*, 666-678. https://doi.org/10.29034/ijmra.v10n1a43

Onwuegbuzie, A. J., & Leech, N. L. (2019). On qualitizing. *International Journal of Multiple Research Approaches, 11*(2), 98-131. https://doi.org/10.29034/ijmra.v11n2editorial2

Onwuegbuzie, A. J., & Leech, N. L. (2021). Qualitizing data. In A. J. Onwuegbuzie & R. B. Johnson (Eds.), *The Routledge reviewer's guide to mixed analysis* (pp. 239-258). Routledge.

Onwuegbuzie, A. J., Ojo, E. O., Burger, A., Crowley, T., Adams, S. P., & Bergsteedt, B. T. (2020). Challenges experienced by students at Stellenbosch University that hinder their ability successfully to learn online during the COVID-19 era: A demographic and spatial analysis. *International Journal of Multiple Research Approaches, 12*(3), 240-281. https://doi.org/10.29034/ijmra.v12n3editorial2

Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351-383). Sage.

Onwuegbuzie, A. J., Witcher, A. E., Collins, K. M. T., Filer, J. D., Wiedmaier, C. D., & Moore, C. W. (2007). Students' perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed methods analysis. *American Educational Research Journal, 44*, 113-160. https://doi.org/10.3102/0002831206298169

Paek, I., & Cai, L. (2014). A comparison of item parameter standard error estimation procedures for unidimensional and multidimensional item response theory modeling. *Educational and Psychological Measurement*, *74*(1), 58-76. https://doi.org/10.1177/0013164413500277

Partchev, I. (2009). 3PL: A useful model with a mild estimation problem. *Measurement: Interdisciplinary Research and Perspectives*, *7*, 94 - 96. https://doi.org/10.1080/15366360903117046

Provalis Research. (2020). WordStat (Version 8.0.28) [Computer software]. Montreal, Quebec, Canada: Author.

Ravand, H., Baghaei, P., & Doebler, P. (2020). Examining parameter invariance in a general diagnostic classification model. *Frontiers in psychology*, *10*, 2930. https://doi.org/10.3389/fpsyg.2019.02930

Reckase, M. D. (2009). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, *21*(1), 25-36. https://doi.org/10.1177/0146621697211002

Reidy, P. (2009). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259-284. https://doi.org/10.1080/01638539809545028

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods, 8*(2), 185-205. https://doi.org/10.1037/1082-989X.8.2.185

Ross, A., & Onwuegbuzie, A. J. (2014). Complexity of quantitative analyses used in mixed research articles published in a flagship mathematics education journal. *International Journal of Multiple Research Approaches, 8*, 63-73. https://doi.org/10.5172/mra.2014.8.1.63

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*(3), 271-282. https://doi.org/10.1177/014662169001400305

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *17*(4), 1-100. https://doi.org/10.1007/BF03372160

Sandelowski, M., Voils, C. I., & Knafl, G. (2009). On quantitizing. *Journal of Mixed Methods Research, 3*(3), 208-222. https://doi.org/10.1177/1558689809334210

Schulz, W., & Fraillon, J. (2011). The analysis of measurement equivalence in international studies using the Rasch model. *Educational Research and Evaluation*, *17*(6), 447-464. https://doi.org/10.1080/13803611.2011.630559

Sébille, V., Hardouin, J. B., Le Néel, T., Kubis, G., Boyer, F., Guillemin, F., & Falissard, B. (2010). Methodological issues regarding power of classical test theory (CTT) and item response theory (IRT)-based approaches for the comparison of patient-reported outcomes in two groups of patients-a simulation study. *BMC medical research methodology*, *10*, 1-10. https://doi.org/10.1186/1471-2288-10-24

Şen, S., & Cohen, A. S. (2023). The impact of sample size and various other factors on estimation of dichotomous mixture IRT models. *Educational and Psychological Measurement*, *83*(3), 520-555. https://doi.org/10.1177/00131644221094325

Seo, D. G., & Kim, J. K. (2021). The accuracy and consistency of mastery for each content domain using the Rasch and deterministic inputs, noisy "and" gate diagnostic classification models: a simulation study and a real-world analysis using data from the Korean Medical Licensing Examination. *Journal of Educational Evaluation for Health Professions, 18*. https://doi.org/10.3352/jeehp.2021.18.15

Sessoms, J., & Henson, R. A. (2018). Applications of diagnostic classification models: A literature review and critical commentary. *Measurement: Interdisciplinary Research and Perspectives*, *16*(1), 1-17. https://doi.org/10.1080/15366367.2018.1435104

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, 8, 1-11. https://doi.org/10.1186/1471-2288-8-33

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583-639. https://doi.org/10.1111/1467-9868.00353

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*, *120*, 103445. https://doi.org/10.1016/j.jvb.2020.103445

Svetina Valdivia, D., & Dai, S. (2024). Number of response categories and sample size requirements in polytomous IRT models. *The Journal of Experimental Education*, *92*(1), 154-185. https://doi.org/10.1080/00220973.2022.2153783

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Applied Social Research Methods Series (Vol. 46). Sage.

Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 3-50). Sage.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative techniques in the social and behavioral sciences.* Sage.

Tharwat, A. (2021). Independent component analysis: An introduction. *Applied Computing and Informatics*, *17*(2), 222-249. https://doi.org/10.1016/j.aci.2018.08.006

Thissen, D., Cai, L., & Bock, R. D. (2011). The nominal categories item response model. In M. W. Van der Linden & C. A. W. Glas (Eds.), *Handbook of polytomous item response theory models* (pp. 43-75). Routledge.

Ultsch, A. (1990). Kohonen's self organizing feature maps for exploratory data analysis. *INNC'90*.

Uyigue, A. V., & Orheruata, M. U. (2019). Test length and sample size for item-difficulty parameter estimation in item response theory. *Journal of Education and Practice*, *10*(30), 72-75. https://doi.org/10.7176/jep/10-30-08

Vehtari, A., & Ojanen, J. (2012). A survey of Bayesian predictive methods for model assessment, selection, and comparison. *Statistics Surveys, 6*(1), 142-228. https://doi.org/10.1214/12-SS102

Viroli, C., & McLachlan, G. J. (2019). Deep Gaussian mixture models. *Statistics and Computing*, *29*, 43-51. https://doi.org/10.1007/s11222-017-9793-z

Waller, N. G., & Feuerstahler, L. (2017). Bayesian modal estimation of the four-parameter item response model in real, realistic, and idealized data sets. *Multivariate Behavioral Research*, *52*(3), 350-370. https://doi.org/10.1080/00273171.2017.1292893

Wattenberg, M., Viégas, F., & Johnson, I. (2016). How to use t-SNE effectively. *Distill*, *1*(10), e2. https://doi.org/10.23915/distill.00002

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA Press.

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, *4*(2), 223-233. https://doi.org/10.1080/15434300701375832