Title

A Guide to the Measurement of Categorical Constructs: A Latent Class Analysis Modeling Approach

Author

Garber, Adam Charles

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

A Guide to the Measurement of Categorical Constructs:

A Latent Class Analysis Modeling Approach

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Education

by

Adam C. Garber

Committee:

Andrew Maul, Chair

Karen Nylund-Gibson

Nancy Collins

December, 2024

The dissertation of Adam C. Garber is approved.

Karen Nylund-Gibson

Nancy Collins

Andrew Maul, Committee Chair

November 2024

Thesis Dedication

From before my admission to graduate school to the very end of my program, I struggled to see myself as belonging in academia. My time in graduate school has been defined by a series of detours and countless missteps. Yet, this decade-long journey of exploration ultimately provided me with a broad and fulfilling education, full of experiences I know will guide me throughout my life. None of this would have been possible without the support and belief from my family, advisors, relatives, and friends, expressed through their care and actions.

This dissertation is dedicated first and foremost to my wife, Haley, and my parents. My mother and father have consistently set the example I strive to follow, and Haley has been my unwavering anchor. Their steadfast encouragement and belief in me kept me on this long and winding path. Without their support, I would have surely lost my way.

I am deeply grateful to my advisors — Mian, Karen, and Andy — who, at different stages of my journey, chose to guide me. Their belief in my potential gave me the external validation I needed to persevere. Each of them, in their unique way, offered crucial wisdom and a steady hand during uncertain times.

The sacrifices my parents have made to support me are immeasurable. Now, as a father to our son Finn, who is 16 weeks old today, I am more aware than ever of the depth of their love and dedication. My parents continue to be my guiding lights, showing me how to live with purpose and selflessness. Becoming a father has given me a new sense of purpose that far surpasses career accolades and ambitions. The joy and responsibility of guiding Finn, and being a present husband to Haley, have become the most meaningful achievements of my life.

Finally, to my partner and closest friend, Haley: you have helped shape me into the person I am today, someone I am proud to be. Your love, patience, and belief in me have made this accomplishment ours to share. You have lifted me up, strengthened my confidence, and made all of this make sense. For that, and so much more, I am eternally grateful.

Adam Garber

October, 2024

VITA OF ADAM C. GARBER

EDUCATION

Doctoral Student (Fall 2024)
Education, Quantitative methods
Interdisciplinary Emphasis: Quantitative Methods in Social Science
Advisors: Andrew Maul, Karen Nylund-Gibson
University of California, Santa Barbara

Master of Arts (Spring, 2017)
Education, Concentration in Special Education & Autism
Advisor: Mian Wang
University of California, Santa Barbara

Bachelor of Arts (Spring, 2011)
University of California, Santa Cruz
Psychology

TEACHING EXPERIENCE

Graduate Associate Teacher (50%; Spring, 2023)
ED 214C: Linear Models for Data Analysis; Dr. Nylund-Gibson

Undergraduate Teaching Assistant (50%; Spring, 2022)
ED 145: Introduction to Research Methods; Dr. Marcoulides

Graduate Teaching Assistant (50%; Spring, 2020)
ED 216C: Structural Equation Modeling; Dr. Nylund-Gibson

Graduate Teaching Assistant (50%; Winter, 2020)
ED 216B: Factor Analysis; Dr. Nylund-Gibson

Graduate Teaching Assistant (Fall, 2019)
ED 201F: Applied Mixture Modeling; Dr. Nylund-Gibson

Graduate Teaching Assistant (Spring, 2019)
ED 214C: Linear Statistical Models for Data Analysis; Dr. Nylund-Gibson

Undergraduate Teaching Assistant (Summer, 2017)
INT 95: Freshman Summer Start Program (FSSP); Dr. Ralph Gallucci

TEACHING WORKSHOPS

Online Course Instructor (Fall, 2022)
QuantFish: Introductory to Advanced MplusAutomation
https://www.goquantfish.com/courses/mplus-automation

Teaching Assistant (Summer, 2022)
Professional Development Intensive Two Day Course: Introduction to Latent Transition Analysis

Methods U, GGSE, University of California, Santa Barbara
Workshop Instructor (Winter, 2021)
Professional Development Two Day Course: Introduction to MplusAutomation
Norwegian University of Science and Technology

Teaching Assistant (Winter, 2019)
Professional Development Mini-Course: Introduction to Latent Transition Analysis
AERA Annual Meeting Toronto, Canada

Teaching Assistant (Summer, 2019)
Professional Development Intensive Two Day Course: Introduction to Latent Transition Analysis.
Methods U, GGSE, University of California, Santa Barbara

ACADEMIC WORK EXPERIENCE

Graduate Student Researcher, MM4DBER (3 quarters; 2023-2024)
Grant: Mixture Modeling for Discipline Based Education Researchers
Position: Gevirtz Graduate School of Education (50% appointment)
Principal Investigator: Dr. Karen Nylund-Gibson

Graduate Student Researcher, IMMERSE (2 quarters; 2022 – 2023)
Grant: Institute of Mixture Modeling for Equity-Oriented Researchers
Position: Gevirtz Graduate School of Education (50% appointment)
Principal Investigator: Dr. Karen Nylund-Gibson

Graduate Student Researcher, Implicit Bias Study (2020 – 2022)
Position: Westmont College (50% appointment)
Principal Investigator: Dr. Carmel Saad

Graduate Statistics Support Peer (3 quarters; 2019 – 2020)
Position: Gevirtz Graduate School of Education (25% appointment)

Graduate Student Researcher, ESTEEM (3 quarters; 2016 – 2017)
Grant: Enhancing Success in Transfer Education for Engineering Majors
Position: Hosted by Department of Chemical Engineering (50% appointment)
Principal Investigator: Dr. Michael Gerber (University of California, Santa Barbara)

GRANT WRITING EXPERIENCE

Grant Writing Research Assistant (2020)
Investigator: Nylund-Gibson

Applied Mixture Modeling Training Workshops for Discipline Based Education Researchers. Submitted to National Science Foundation (NSF).

Grant Writing Research Assistant (2019)
Investigator: Nylund-Gibson
Applied Mixture Modeling Training Workshops and Resources for Education Researchers. Submitted to Institute of Education Sciences (IES).

Grant Writing Research Assistant (2015)
Investigator: Wang & Singer
A comparative study of patterns of family support for children with developmental disabilities:
Similarities and differences in U.S. and China. Submitted to Department of Education.

PUBLICATIONS

Nylund-Gibson, K., Garber, A. C., Carter, D. B., Chan, M., Arch, D. A. N., Simon, O., Whaling, K., Tartt, E., & Lawrie, S. I. (2022). Ten frequently asked questions about latent transition analysis. Psychological Methods. Advance online publication. https://doi.org/10.1037/met0000486

Nylund-Gibson, K., Garber, A. C., Singh, J., Witkow, M. R., Nishina, A., & Bellmore, A. (2023). The Utility of Latent Class Analysis to Understand Heterogeneity in Youth Coping Strategies: A Methodological Introduction. Behavioral Disorders, 48(2), 106–120. https://doi.org/10.1177/01987429211067214

Denson, N., Ing, M., Arch, D., Garber, A. C., Chan, M., Carter, D., Nylund-Gibson, K. (2022). Interactions with Diversity in Higher Education: A Latent Class Analysis. Journal of Diversity in Higher Education. Advance online publication. https://doi.org/10.1037/dhe0000451

Felix ED, Afifi TD, Horan SM, Meskunas H, Garber A. (2020) Why Family Communication Matters: the Role of Co-rumination and Topic Avoidance in Understanding Post-Disaster Mental Health. Journal Abnormal Child Psychol. 48(11):1511-1524. https://doi:10.1007/s10802-020-00688-7

PRESENTATIONS

Garber, A. & Nylund-Gibson, K. L. (2021) Structural Invariance in Multigroup Latent Class Analysis: Perception of Disability Status and Academic Expectations [Symposium]. AERA Annual Meeting (virtual)

Garber, A., Wang, M. & Nylund-Gibson, K. L. (2021) Typologies of Autism: The Utility of Latent Class Analysis to Understand Heterogeneity in Social Responsiveness [Symposium]. AERA Annual Meeting (virtual)

Garber, A. & Nylund-Gibson, K. L. (2020) Structural Invariance in Multigroup Latent Class Analysis: Perception of Disability Status and Academic Expectations. GGSE Interdisciplinary Research Symposium Santa Barbara, CA.

Nylund-Gibson, K. L. & Garber, A. (2020) Understanding Change in Latent Transition Analysis Models Using Auxiliary Variables [Symposium]. AERA Annual Meeting San Francisco, CA http://tinyurl.com/vuslpat

Carter, D., Garber, A., Nylund-Gibson, K. L., Dowdy, E. & Furlong, M. J. (2020) Understanding Patterns of Social-Emotional Strengths Across Students With and Without Disabilities: Different, Not Worse [Symposium]. AERA Annual Meeting San Francisco, CA http://tinyurl.com/qlyypv5

Garber, A. & Nylund-Gibson, K. L. (2019) Conditional Mediation in Latent Class Analysis [Symposium]. AERA Annual Meeting Toronto, Canada

Garber, A. & Wang, M. (2017) Responsive Early Education and Intervention: A Call for School and Family Collaboration [Poster]. CEC Annual Meeting Boston, MA

WORK EXPERIENCE

Behavior Interventionist (2016 – 2020)
Koegel Autism Center
University of California, Santa Barbara

Behavior Interventionist (2014 – 2017)
Holdsambeck & Associates
Santa Barbara, CA

Behavior Interventionist (2014)
Trumpet Behavioral Health
Santa Barbara, CA

Mental Health Worker (2010-2011)
7th Avenue Center: Mental Health Psychiatric Hospital
Santa Cruz, CA

Abstract

A Guide to the Measurement of Categorical Constructs:

A Latent Class Analysis Modeling Approach

by

Adam C. Garber

In this thesis I present a guide to constructing a categorical measure to conform to the unique structure of a latent class nominal variable. A series of detailed steps are outlined from the starting point of construct definition through to the evaluation of models with construct validity as a common focal point at each stage of construct development and evaluation. The process of defining a construct to measure a categorical property and evaluate its validity is referred to in this thesis comprehensively as the practice of *constructing measures*.

Currently, a pedagogical study which provides guidelines to applied education researchers for how to design a categorical construct, evaluate measurement, and build evidence for construct validity remains unaddressed in the literature. Furthermore, few applications of latent class analysis (LCA) exist which describe a process of developing a categorical construct from a measurement-oriented perspective. It is this author's hope that by providing guidelines for constructing categorical measures researchers applying LCA methods will utilize these measurement practices and emphasize construct validity in their own research. Considering the construction of a categorical measure from the ground up requires a novel orientation towards measurement to be taken. Starting a discussion about the

unique challenges of constructing a categorical measure provides an important contribution to the social sciences literature.

This dissertation will provide a roadmap for defining and measuring a latent categorical construct. To demonstrate how construct measurement may be applied in practice a series of examples from education will be incorporated throughout the thesis. This paper is targeted toward applied education researchers with task-oriented recommendations intended to provide a starting point for researchers to engage in the practice of constructing measures.

Table of Contents

Chapter 1: Introduction

Statement of problem to address

In the field of educational measurement, well-defined procedures have been developed to guide the construction of instruments used to measure the attributes of students, teachers, classrooms, and schools. However, in this literature limited landscape has been dedicated to the measurement and epistemology of categorical constructs. This paper aims to initiate a discussion about how categorical properties may be measured and constructed. The popular use of LCA models without clear guidelines for constructing categorical measures results in ambiguity surrounding the semantics, epistemology, and ontology of such constructs.

Applications of LCA in education and the social sciences commonly begin with the utilization of previously existing scales or secondary data (see review; Appendix A). Often, these repurposed scales (Slaney & Racine, 2013) were not originally designed to measure categorical properties. Consequently, issues of construct definition are frequently underdetermined or only briefly addressed in publication. In a review conducted for this thesis, the most common method used for developing an LCA construct involved a process of selecting a subset of items from a pre-existing scale (Appendix A). Additionally, many of these studies re-code item responses by conflating or averaging response options. Among the 100 reviewed studies applying LCA methods to education data, 81% used existing scales, 63% used secondary data, and 57% re-coded item responses (Appendix A). Notably, only four of the applied studies from the review utilized confirmatory LCA methods to test theories of categorical constructs (Gao et al., 2020; Jansen & van der Maas, 1997). In summary, a clear opportunity exists to study how measurement practices can be extended to inform the study of categorical constructs.

Irrespective of the initial approach taken in an LCA analysis, researchers can benefit from considering how categorical constructs are measured and how to refine constructs using theory-driven methods. This thesis is intended to encourage contemplation about what makes a categorical construct valid. Validity is a pertinent consideration at any research stage and can be incorporated into research utilizing secondary data as well as those embarking on the development of an LCA construct from scratch. Researchers may view these guidelines as a menu of recommended methods, acknowledging that constructing measures constitutes a non-linear, iterative process, amenable to diverse methodological perspectives.

Initiating a discussion about the measurement of categorical constructs can offer several advantages for research programs which apply LCA models in education. First, by providing a detailed guide to how latent categorical measures are constructed, procedures may be established and then subjected to refinement as the field progresses. Second, prioritizing construct validity in LCA studies may incentivize researchers to allocate resources towards assessing the replicability of existing LCA constructs (Schmiege et al., 2018). Third, by highlighting how specific research objectives shape the definition of latent categorical constructs, the connection between theoretical precedent and analytic practice within LCA applications may be strengthened. Lastly, by integrating the aforementioned practices, research norms could evolve to prompt applied LCA researchers to more extensively discuss issues of construct validity in publication settings.

Thesis objectives

The objective of this dissertation is to delineate a procedural framework comprising actionable steps for researchers to construct a categorical measure. This paper examines the following practical tasks faced when constructing a measure, beginning with defining a categorical construct and progressing to instrument evaluation utilizing the LCA model framework. The process of constructing a categorical measure can be broken down into six broad measurement tasks (Figure 1):

1. Statement of research goals and motivations: Develop specific research goals and transparently state motivations for the study.

2. Conceptual development: Follow the iterative steps of categorical construct map procedure.

3. Item design: Write a series of candidate survey questions

4. Construct revision and item selection: Re-evaluate the construct, revise construct definition as needed, and select a subset of items to constitute the LCA indicator set.

5. Post-data analysis: Examine the data and assess the alignment between the model and the construct theory.

6. Replication: Validate the findings through replication.

These steps will be demonstrated using an applied example of a categorical construct proposed to measure a property constituted by typologies of *adaptive flexibility* for adults with autism. This example aims to demonstrate how each measurement task translates to research practice, particularly within the applied context of measuring heterogeneity in adaptive flexibility among populations of adults with autism.

Figure 1

Diagram Outlining the Steps of Constructing Categorical Measures

Constructing Measures							
Construct Definition				Measurement			
State: Research questions	Propose: Initial construct definition	Refine: Construct map	State: Updated construct definition	Item design	Construct revision & item selection	Model evaluation	Replication, revaluation, and validation

Note: The phases of constructing measures may follow an iterative or recursive process.

Pedagogical approach

Theory-first. This dissertation advocates for a theory-first approach, positing that construct definition should precede the selection of a statistical model. From this standpoint, during the conceptual phase of construct development, constructs are described independently of any specific statistical models (e.g., latent class constructs). Although a categorical construct may subsequently be appropriate to study using LCA methods, in this thesis, the language *categorical construct* is used to refer to a construct in the conceptual phase of the measurement process to emphasize that model selection should be driven by theoretical considerations. This viewpoint underscores the fundamental notion that the initiation and structuring of a research endeavor ought to be carefully articulated and preceded by theoretical grounding (Borsboom et al., 2004), prior to the contemplation of quantitative analysis techniques.

Defining terms. To clarify the intended meaning of the language used in this thesis it is important to provide definitions of some commonly used terminology. Terms, such as the word *construct*, are used broadly across research programs and applied settings and adopt nuanced differences in definition which underscores the need for precise and consistent definitions (Slaney, 2017). In the following section, we provide a taxonomy of important terms associated with the development of a categorical construct and the measurement of a property or attribute (see Table 1). The following terms have been defined intentionally to provide a framework for constructing measures tailored to the assessment of categorical properties.

In the following discussion, the term *construct* is used to refer to the theoretical concept being studied and is distinguished from the *property* or *attribute* believed to exist which is the target of measurement (e.g., participants have *attributes* and researchers *construct* theories). This implies that constructs are ideas developed by researchers and by definition abstractly represent or describe a focal property or set of attributes (Slaney & Racine, 2013). In contrast, the terms *property* and *attribute* are used to refer to the object of study (i.e., the referent), or the phenomenon that exists in the population of interest that we wish to measure. In education and the social sciences properties of interest are often not directly observable (e.g., adaptive flexibility), hence the need to describe such properties conceptually and create models to study the properties effects in the population of interest. In the case of a self-report scale a property is assumed to be the cause of the item responses or the mechanism which (at least in part) determines the item response process for participants in the sample.

The term *model* is defined here, in a general sense, as referring to a quantitative representation used to relate observed items (i.e., variables) to the construct of interest. In this thesis, the term model is used most frequently to refer to a latent variable model, specifically the LCA model, and describes the quantitative relation between the item responses and the latent class variable. The term *model* is used both to reference the mathematical specification of the latent variable model as well as the theoretical assumptions implied by the formal structure of the model (i.e., statistical assumptions). Consequently, the particular model used by a researcher represents a deliberate choice constituting a theoretical assumption about how the observed item responses relate to a construct.

The term *measurement* is used in this thesis to describe the empirical process of taking a set of survey items (i.e., inputs) and using this information to make inferences about a property described by a model of a latent construct (e.g., latent classes; Mari et al., 2021). The process of measurement for a categorical construct modeled using the LCA approach can be described using four empirical steps: 1) Define the categorical construct based on the property of the population you wish to study, 2) collect observations using a survey of items, 3) model the relations between the observed scores and the measurand using a categorical latent class model, and 4) utilize the results of the latent class analysis to make inferences about the property of interest. Each of these steps necessitates theory-driven decisions including how the construct is defined, how the measurement is conducted, how the model is specified, and the inferences derived from the LCA results.

Table 1

Defining Terms: Construct, Property, Model, Categorical Measurement

Term	Definition	Applied Example
Construct	A theoretical concept constructed by researchers to abstractly represent the focal property or attribute	*The adaptive flexibility construct refers to a series of behavioral attributes for populations of adults with autism*

Property, Attribute	The phenomenon that exists in the population of interest that we wish to measure (i.e., the referent)	*Behavioral attributes of populations of adults with autism (e.g., social flexibility)*
Model	A quantitative representation used to relate observed items to the construct of interest	*A three class LCA model including dichotomous items and a nominal latent variable*
Categorical measurement	An empirical process used to make inferences about a population attribute which relates observations to a model of a categorical construct	*Adaptive flexibility survey items are used to make inferences about behavioral attributes using LCA to model a categorical construct*

Choosing an approach to constructing measures. This thesis posits that clearly defining a categorical construct based on theoretical precedent provides the strongest evidence to support the validity of a measurement instrument. The guidelines proposed may be viewed as outlining a series of suggested best practices for constructing measures given the explicit goal of establishing evidence for the validity of the construct. However, it is recognized that researchers may have alternative motives for utilizing LCA methods and therefore may find some of the recommended practices more or less applicable to their own research agendas. In this case, it is advisable that readers selectively engage with sections pertinent to their research interests and modify the suggested procedures to align with their research purposes. The qualitative procedures for construct development that are proposed involve engagement with the population under study which is often a time-intensive enterprise. Acknowledging that realized economic constraints vary broadly by academic context, researchers are encouraged to choose which practices are feasible to implement in their respective research settings.

As a divergence from the conventional practice of LCA, which has been dominated by exploratory methods, this thesis proposes an alternative approach to implementing an LCA analysis beginning with constructing a categorical measure from the ground up in a theory-driven confirmatory manner. The process proposed begins with the qualitative task of defining the construct and then progresses to examining the structure of a categorical attribute through confirmatory latent class analysis. However, given that confirmatory practices remain relatively uncommon in applied LCA research (see Appendix B), researchers may find some combination of confirmatory and exploratory strategies to provide a reasonable compromise to the measurement-oriented practices recommended in this thesis. For example, researchers might adopt a hybrid approach, hypothesizing certain elements of the latent class variable structure, such as the number of classes, while exploring others, like the form and meaning of classes or their sizes, through exploratory methods.

Alternative approaches, such as the use of hybrid confirmatory and exploratory research strategies will be outlined and contrasted with the confirmatory approach. The underlying premise is that any incorporation of theory-driven practices represents a constructive step towards enhancing the quality of measurement and the validation of constructs in LCA research. Furthermore, researchers who traditionally utilize exploratory LCA methods, may find this discussion beneficial for formulating research questions, defining constructs, and addressing questions of construct validity.

Target audience. This thesis is designed to serve as a pedagogical resource for applied education researchers, particularly graduate students engaged in statistical training within education and social science disciplines who may be considering the measurement of a categorical construct for the first time. It aims to bridge the disciplinary divide between measurement theory and statistical methodologies by describing the intersection of these technical subjects in an accessible way. Moreover, the effectiveness of this thesis in promoting advanced measurement practices among students is contingent upon its acceptance and endorsement by methodologists who are responsible for teaching LCA techniques. To this end, space is dedicated to persuading methodologists of the critical need to integrate measurement practices into the curriculum of quantitative methods programs. This thesis may also be of interest to special education researchers, as it includes an applied example examining the construct adaptive flexibility among adults with autism, thereby illustrating the concepts discussed.

Introducing the Applied Example – The Adaptive Flexibility Construct

This thesis utilizes an ongoing example to demonstrate the efficacy of the outlined procedures for constructing measures within an applied context. The construct chosen, named *adaptive flexibility*, is proposed to measure a property of adult populations with autism. This construct is defined by typologies of adaptive flexibility, a concept building off previous work in special education and psychology (Didden et al., 2008). It aims to measure the presence of flexibility or inflexibility across various behavioral domains, identified as critical to adaptive functioning by previous studies in special education (Green et al., 2006) and autism research (Kensworthy, 2008). Specifically, a theory of qualitatively distinct patterns of adaptive flexibility is evaluated by engaging with the categorical measurement procedure outlined in this thesis. In the example, a sample of qualitative data is collected to illustrate item development and the validation process. Furthermore, a panel of specialists in autism, psychology, and measurement fields are consulted to inform construct definition and item design.

The pedagogical demonstration of the LCA analysis utilizes simulated data. Features such as the latent class form and class size used for the adaptive flexibility example are informed by conceptual theory and data is simulated to illustrate contexts of agreement and disagreement with this substantive theory. It should be noted that the data used for the LCA analysis are simulated for illustrative purposes only, and any conclusions about the subject population would necessitate empirical validation in future research endeavors.

In addition to the primary example of adaptive flexibility, this thesis will incorporate supplementary examples for pedagogical purposes. These examples will illustrate the application of construct development and measurement concepts across a variety of applied social science research contexts. Such supporting examples, drawn from a sample of reviewed articles listed in Appendix A, will elucidate how contextual factors influence decisions related to measurement approaches, construct definitions, and modeling assumptions. The primary and supportive examples are intended to connect theoretical concepts with practicable illustrations across a range measurement contexts commonly encountered in education. Taking an intentionally pragmatic position towards measurement, conceptual material presented will be accompanied by examples and action-oriented recommendations. This position entails tying construct development and measurement concepts closely with practicable solutions.

Steps for constructing categorical measures

The constructing measures process presented here constitutes a system or collection of methods for conceptually defining, refining, and evaluating the validity of a categorical construct. The approach to constructing measures proposed is organized into five general stages as outlined below. A reference point formative in structuring the methods proposed in this thesis is Wilson's (2005) "four building blocks" approach and the concept of the construct map. However, procedurally the constructing measures method presented diverges significantly in content from this reference point as categorical constructs require the reconceptualization of methods designed to measure continuous properties.

Part 1: Development of study goals and motives. The first step of constructing measures proposed is to identify common research goals which motivate the study of categorical properties. The purpose of this section is to articulate how specific research motives influence construct development and dictate which methods are most appropriate to evaluate or model such constructs. Here it is assumed that the inferential strategy best suited to study a categorical construct will vary based on the topical area, questions, motives, and perspectives of the researcher. Specifically, substantive factors such as characteristics of the population of interest and the particular attributes of that population that a researcher chooses to study should be considered when choosing an inferential strategy (Wilson, 2003). Furthermore, a researcher's background and training will play a significant role in determining which inferential strategy is considered. Consequently, a one-size-fits-all approach to constructing measures is not endorsed in this thesis. Additionally, it is argued that the strategy used to define, measure, and model categorical constructs should be tailored to fit the theoretical orientation and goals of the researcher. After research questions have been clearly and precisely proposed the work of defining a categorical construct can commence.

Part 2: Construct definition and the categorical construct map. This section focuses on the need to clearly define research goals and construct definitions, arguing that these steps are essential for guiding analytical decisions and to build a case for construct validity. The inferential goal of measuring categorical properties differs fundamentally from the measurement context underlying Wilson's (2005) construct map, which focuses on the investigation of continuous properties. As such, the construct map concept in its original form is not directly applicable to categorical constructs. To address this, an alternative conceptualization of the construct map, specifically tailored to the development of categorical constructs, is proposed. This approach involves the production of a sequence of tables and figures, providing a systematic method for defining a categorical construct.

To describe a categorical construct, it is useful at this time to introduce the concept of *domains*– sub-construct elements that together form the categorical construct. In this thesis, domains are considered the theoretical precursors to items, describing the components of the construct before the formal development of an instrument. A detailed discussion of how decisions regarding domains, such as inter-domain relationships, influence construct definition can be found in Chapter 2.

The proposed categorical construct map is an iterative procedure used to identify candidate domains, refine the construct based on new information, visually represent potential class patterns, and evaluate the construct comprehensively. The phases of the construct map procedure are summarized in Table 2. The phases outlined in the table, though organized in sequence, are flexible and multiple cycles of refinement may be needed to define the construct.

Table 2

Construct map Phases for Constructing Categorical Measures

Phase	Title	Description
Phase 1	Draft Construct Map	List profile patterns based on practitioner observations. Starting with a minimal case is recommended (i.e., 2-3 domains).
Phase 2	Expand Construct Map	Add domain (columns) to the map. Goal is to make construct coverage comprehensive in-line with theoretical bounds. Each domain can be seen as a candidate item (proceeding the item design phase)
Phase 3	Refine Construct Map	Cut redundant or highly overlapping domains. Columns with high content-overlap can be identified by strong correspondence with other domains on the map. Note– items may continue to be refined (cut) at a later point
Phase 4	Test & Revise Construct Map	An iterative process of pilot testing, revising, and updating the construct map. Steps 2-3 may require repeating until construct refinement is complete

Part 3: Item design. Proceeding the construct definition phase, the next task of construct design is to generate survey questions which align with the substantive domains chosen in the last iteration of the construct map. A series of steps for item design are

proposed to draft an initial set of survey questions and build evidence for the content validity of the items based on feedback from the study population and relevant stakeholders. The perspective of item development endorsed in this paper is that the respondent population is the principal resource for providing insight about the cognitive response process realized by respondents when engaging with a survey instrument. Consequently, a series of qualitative techniques developed to evaluate survey validity based on the item response process (Maul, 2018; Wolf et al., 2021) are recommended and considered with respect to the measurement of a categorical property. The methods proposed are aligned with the following goal of item design– to construct items which consistently and accurately measure the attributes of interest in the target population. Furthermore, by collaboratively evaluating survey instruments with the participation of the community under study this approach provides agency to the study population which aligns with the ethical imperative to conduct culturally sensitive research (Balcazar et al., 1998).

Part 4: Construct revision and item selection. In this chapter a set of procedures are outlined to evaluate construct scope and determine indicator assignment based on considerations of construct coverage and sub-construct balance across respective domain areas. Once a set of survey questions have been composed, a focal shift towards evaluating the construct as a whole is useful to assess whether the content area covered by each item fits within the theoretical bounds defined in the construct map phase and aligns with the research aims proposed to motivate the study. At this point, research goals and construct definitions stated earlier in the constructing measures process may warrant revision to accommodate information acquired through engagement with the substantive topic and the study population. The goal of this phase of constructing measures is to calibrate survey items with construct theory and to determine which items are best suited to constitute indicators of the LCA model. The end of the construct revision and evaluation process marks a departure point in this thesis from the presentation of qualitative strategies for construct development to post-data measurement strategies for building quantitative evidence and testing models to make inferences about categorical properties.

Part 5: Post data collection construct refinement. In this guide to constructing categorical measures, the continuation of theory-driven practices is advocated to lead the course of analysis after data collection. Following a construct-focused perspective, guidelines are outlined for the utilization of data to evaluate construct definitions and conduct confirmatory analyses using the LCA model. Initially, a descriptive technique is recommended to assess response patterns by visually representing the large contingency space created by the joint distribution of the indicators. This response pattern information can be used to confirm or disconfirm prior hypotheses about the construct as well as inform model specification decisions. Subsequently, inferences about the measurement of a categorical property are evaluated following the methodological assumptions posited by latent variable theory. Several confirmatory LCA approaches are outlined, enabling researchers across various substantive contexts to select an analytical procedure with varying degrees of confirmatory strength. Confirmatory LCA (CLCA) approaches are argued to be effective at building an evidentiary case for the validity of a categorical construct. Taking a long-view of the developmental horizon of a construct, the validation of a construct may be seen as a process spanning multiple studies that benefit from coordinated research efforts (Fink et al., 2021). In light of this perspective, hybridized exploratory-confirmatory approaches are considered with focus concentrated on optimizing the long-range replicability of the construct findings. The final section of this chapter will explore how prioritizing specific inferential goals influences research practices in the study of categorical constructs.

Part 6: Replication. An important method for evaluating the validity of a categorical construct is to replicate LCA results across independent samples from a population (Finch & Bronk, 2011). Confirmatory LCA methods provide a means to build evidence that the internal structure of a categorical construct can be consistently identified in the focal population. Replication of a study results can provide evidence that a categorical property is measured reliably and reinforce the case for the construct's validity. Although a small body of applied LCA replication studies have been conducted (e.g., Laudy et al., 2005; Gerber et al., 2009), increased replication efforts would provide a valuable complement to the primarily exploratory research found in the applied LCA literature. In this thesis guidelines are provided to increase the accessibility of replication methods and raise awareness about how replication can strengthen the validity of theories which propose to measure categorical properties.

Background literature

Measurement. The call for increased focus on construct definition and construct development practices in the social sciences is a topic that has been widely vocalized by researchers in psychometric and measurement fields (Maul, 2017; AERA et al., 2014). The circumvention of measurement issues in applied research is not specific to studies which utilize LCA methods but rather is an issue prevalent across applications of quantitative methods in the social sciences (Flake et al., 2017; Borsboom, 2008). Literature which addresses measurement issues for categorical latent variables specifically is sparse, including few texts (i.e., Borsboom, 2005) and isolated studies on specialized topics (e.g., item sensitivity; Cole et al., 2017). However, many concepts from the general measurement and validity theory literature can be applied or adapted to inform procedures for constructing a categorical measure. The extensive body of work developed by measurement scientists offers a strong foundation for addressing the question of how to go about the measurement of a categorical property. A central aim of this thesis is to provide a non-technical discussion of measurement and validity issues in an approachable format for applied social scientists. Moreover, the author intends to highlight practical construct development procedures prominently in the discourse, ensuring that applied researchers are first acquainted with the goal-oriented purposes of these practices before delving into the more theory-intensive aspects of measurement and validity.

The term *measurement* is used in this paper to refer to the empirical process of relating observed variables to a latent class variable to draw inferences about a focal categorical attribute. Although what is theorized to constitute *measurement* continues to be a contentious issue (e.g., Michell, 2012), there have been several recent calls advocating for expanding measurement definitions to include classificatory attributes (Mari et al., 2017). Assuming the study of all constructs, continuous and categorical alike, can benefit from the collection of techniques identified as measurement and construct validation practices, this seems to be an easily defensible position.

Within the context of LCA, measurement involves using observed response patterns to sort populations into distinct latent classes, thereby facilitating inferences about categorical properties.The concept of measurement will therefore be utilized here to describe how meaning is ascribed to nominal latent variables including what makes classes distinct in relation to the other classes, the conceptual salience of each class individually, and the construct coherence holistically. Given the prevalent use of classification, and categorical groupings used to determine student outcomes in education, there is a clear and pressing need to develop robust methods for measuring and validating categorical constructs.

Validity. This thesis proposes that the principles of validity established in the general psychometric literature (AERA et al., 2014) can guide the measurement of categorical properties. Methods for collecting validity evidence (i.e., validation) proposed in the *Standards for Educational and Psychological Testing* (AERA et al., 2014) are argued to generalize to the special context of the measurement of categorical attributes. Although various definitions of validity exist (Markus & Borsboom, 2013), this work emphasizes that practices for collecting validity evidence should be contingent upon the construct's intended purposes and the realized consequences of its use (Maul, 2018). Essentially, what constitutes validity evidence depends on the purposes motivating the studies research questions and the ethical imperatives towards the population under study. In the applied LCA literature reviewed (Appendix B), a significant portion of studies were identified where construct definitions were underdetermined or required further clarification. A critical piece of the validity argument is to articulate how findings from latent class analyses should be interpreted and to clearly specify the contexts in which these findings are and are not applicable. This places the responsibility on the researcher to transparently delineate the scope of the construct by defining its meaning, discussing what does or does not constitute the construct, and stating where and how it should be used.

The validation process entails gathering evidence to support the proposed use and interpretation of a construct, along with determining the relevance of different types of evidence to the specific research context. This thesis discusses several types of validity evidence for categorical constructs, including: 1) Evidence of alignment between construct definitions and the realized scale content; 2) Expert judgements of construct representation in terms of quality, comprehensiveness, and construct-irrelevance; 3) Insights from respondent populations on their cognitive response process when answering scale items; 4) Evidence based on relations between categorical constructs and auxiliary variables (AERA et al., 2014). A key contribution of this paper is to highlight the significance of addressing validity issues specific to LCA research contexts and to propose specific, actionable procedures for collecting evidence for validity.

Education. This thesis is further informed by the body of research methods, applied theories, and measurement practices developed in the field of education. The literature reviewed addresses specific challenges of categorical measurement as it applies to the study of students, teachers, classrooms, and schools. Researchers in education have explored a wide array of categorical constructs using the LCA approach, investigating diverse topics such as typologies of dual language learners, preschool teachers, student aptitudes, and school environments (Kim et al., 2018; Nasiopoulou et al., 2107 ; Gao et al., 2020; Duff & Bowers, 2022). The guidelines proposed aim to address a range of prototypical substantive issues encountered in educational research. This objective is supported by the primary and supplementary examples discussed previously. The extensive body of applied LCA research serves as a critical reference for determining which types of constructs are viable candidates for modeling within the latent class framework.

Introduction to the LCA measurement model

In employing the Latent Class Analysis (LCA) approach to model a categorical construct, the form of the model carries theoretical implications that influence how meaning is attributed to the construct. Many pedagogical resources are available which describe the statistical features of the LCA model and are written to be accessible to applied researchers in education and social science (Nylund-Gibson & Choi, 2018; Masyn, 2013; Hagenaars & McCutcheon, 2002). A review of the general features of the LCA model is warranted before procedures for constructing measures can be proposed. In this thesis, I provide an overview of the LCA model with emphasis on characteristics of the model which are integral to how a latent categorical construct is defined and interpreted. While various extensions of the LCA model exist, this thesis focuses on the most common application of the LCA model which includes dichotomous indicators, unordered classes, and assumes conditional independence. The statistical assumptions of the LCA model are covered in the following section. The adaptive flexibility construct example is used to illustrate the relationship between a researcher's theoretical assumptions and the constraints imposed by the structural features of the LCA model.

Situating LCA in the latent variable framework. LCA is a type of model from a broader family of models referred to conventionally as latent variable models. Common types of latent variable models used in education research include factor analysis (FA), item response theory (IRT), and LCA, among others (Muthén, 2001). The latent property of these models refers to measurement of an unobserved construct which is not directly but indirectly measured by one or more observed items. A feature of FA and IRT models, key to their identification, is the inclusion of a latent variable that is modeled as a continuous property. In contrast, LCA and other mixture models incorporate a latent variable modeled as a categorical property, making them particularly suited for the analysis of categorical constructs. Therefore, among researchers who ascribe to the latent variable measurement perspective and have proposed a theory about categorical constructs, mixture models like LCA provide an analytical approach explicitly qualified for the study of categorical constructs.

The LCA model is one member of a family of models called mixture models. Mixture models have in common that they describe the joint distribution of a variable set by sorting response patterns into a finite set of composite distributions (i.e., categories). This mixture of latent categories is commonly referred to in the literature as groups, classes, or profiles. In this thesis we focus on one specific mixture model latent class analysis (LCA), which is among the least complex in the model family. For consistency and clarity, in the context of the LCA models, I will refer to latent categories as *classes*.

Measurement models and observed measures. The LCA measurement model serves as the means by which observed response patterns are used to describe or summarize the construct. In essence, this model is a statistical tool employed to infer meaning about the construct of interest from a sample of the respondent population. It is important to distinguish that the measurement tasks that precede data analysis, which involve construct definition and instrument development, are a related but distinct process from the *measurement* referred to in the discussion of the *LCA measurement model*[1]. The former involves conceptual tasks for defining a construct, while the latter assesses whether an instrument is adequately measuring a property based on constraints (assumptions) imposed by the measurement model. The distinction of the LCA model as a *measurement model*, as opposed to just a model, lies in its application: this modeling approach is used to measure an unobserved categorical latent construct. Specifically, how the observed response patterns are assumed to be measures of latent class membership and determine the categorical form of the construct.

Model assumptions. In the following section, I will outline a series of modeling assumptions pertinent to the LCA model context. Each assumption about the model inherently relates to an assumption about measurement, carrying specific substantive or theoretical implications. For instance, the conditional independence assumption, while seemingly abstract, significantly influences how we interpret analytical results and constrains the scope of permissible inferences. Throughout this thesis, we will continuously evaluate whether the assumptions underpinning the LCA model align with or contradict the theoretical assumptions deemed appropriate in our field of study. In certain instances, the modeling assumptions will naturally align with our theoretical understanding of how the construct should describe the population under study. Conversely, a modeling assumption may have no clear justification, or be at odds with existing theory, which may in-turn weaken the validity of our inferences. The tradeoffs faced when making common LCA modeling assumptions will be demonstrated using the ongoing example as a means to explain the

[1] Note that this semantic distinction is made to follow the convention in the mixture modeling discipline. It is argued in some measurement areas that any statistical model by definition is functioning to achieve measurement in some capacity. Making the distinction of 'measurement model' redundant.

statistical features of the LCA model in a practical and intuitive manner. By assessing the implications of LCA model assumptions in specific research contexts, this thesis aims to assist practicing methodologists in identifying and transparently addressing areas of discord or agreement between the measurement model and construct theory. An assumption central to the frequentist statistical philosophy is that the specified model is the correct model; all other modeling assumptions in the following discussion are predicated on this principal question.

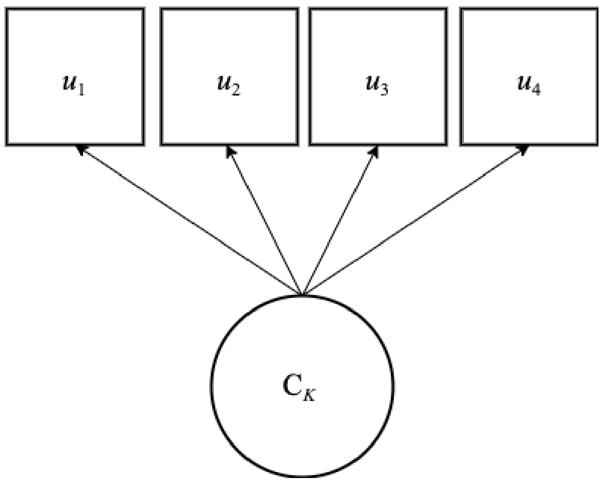
Model under-determination. In the study of human attributes which use quantitative methods to classify populations into groups it is particularly pertinent to acknowledge that *models* are by definition abstractions from reality (MacCallum, 2003). The abstract nature of latent variable models is highlighted by the statistical phenomenon known as model equivalence which illustrates that for any joint distribution of data multiple latent variable models will fit the data equally (MacCallum et al., 1993). This issue is explained by the concept of model under-determination and implies that model fit information alone cannot provide conclusive evidence that a particular model is correct. Multiple sources can contribute to the under-determination of our inferences (i.e., lacking accuracy or precision), such as inadequate construct definitions, imprecise instruments, or incorrect model assumptions. However, it is important to distinguish here that latent variable models are by definition statistically under-determined (Borsboom et al., 2003), highlighting the necessity for theory to take precedence in guiding modeling decisions. This principle is critical in all aspects of the LCA model; every model assumption, including the choice to measure the construct as categorical, the selection of items to measure the construct, and the configuration of the latent variable (i.e., number of classes, conditional independence), must be justified theoretically.

Indicators. In LCA the categorical construct is assumed to be measured by a discrete set of observed items. Once these items have been administered to the respondent population and the responses have been collected the resulting variables used to measure the latent class variable are referred as *indicators*. The term *indicator* describes how the items are modeled to relate to the latent variable, specifically as indicators of latent class membership. Considering the adaptive flexibility construct example, the item set chosen to measure aspects of adaptive flexibility, after survey administration, are then utilized as indicators of the LCA measurement model. The specified model is then used to make inferences about the adaptive flexibility construct's validity as a categorical measure of latent class. The LCA indicator set jointly provides the information which will determine the structure of the latent categorical constructs form including the number of classes, class size, and class shape. The fact that the latent class variable is measured exclusively by the indicators selected by the researcher, underscores the importance of carefully choosing the final indicator set which will in-turn determine the structure and meaning of the latent classes.

The reflective model. In LCA, the nominal latent variable attains the status of the *measurand* (i.e., the property being measured), hence the nomenclature of referring to latent variable models as *measurement models*. As shown in the path diagram, the LCA model indicators are specified as dependent variables with the regression arrows pointing from the latent variable towards the observed indicators ($C_k \rightarrow u_1 \ldots u_i$; Figure 2). This directionality of the regression arrows represents an assumption of the LCA measurement model as it implies that the latent class variable is the antecedent cause of the response patterns observed by the indicator variables. This specification arrangement suggests that the latent class variable is the common cause of the indicator variables and is described in psychometrics as a *reflective model* (Borsboom; p. 61).

Figure 2

The Reflective LCA Model



Returning to our example, the LCA model implies that the adaptive flexibility class typologies are responsible for the response patterns observed in the sample population. In other words, it is because a respondent has the attribute of socially adaptive flexibility that they respond a certain way to the survey questions. To better understand the implications of the reflective model we may consider what would happen if the arrows in Figure 2 were reversed ($x_1 \ldots x_i \rightarrow C_k$). This model is called the formative model and implies that the latent variable represents a summative metric composed by various independent properties (e.g., SES). In the formative model the observed indicators attain the status of the properties of measurement themselves, hence no measurement is being modeled and therefore the model is not given the classification as a *measurement model*.

Applying the formative model logic to the ongoing example suggests that the sampled population have endorsed multiple socially related indicators that can be described as a socially adaptive typology representing a composite of properties. Notice the distinction in the ontological meaning ascribed to the latent classes: for the reflective model the adult responds to the survey question in a particular manner because they possess the attribute socially adaptive flexibility; conversely, in the formative model each response is influenced by their predisposition toward each question, with the latent variable representing a composite classification of these predispositions—the cumulative effect of a series of causes.

The reflective and formative models suggest different ontological perspectives regarding the status of the latent variable. In this thesis, I take a neutral stance as to how literally researchers should interpret the causal implications of this modeling feature. For an in-depth discussion of this topic, see Borsboom (2005).

Conditional independence. In the LCA model, the assumption of conditional independence plays a central role in how we understand and interpret latent classes. This assumption implies that after conditioning on latent class, the covariation in indicator responses is fully explained. This means that after conditioning on the latent classes the indicators are assumed to be independent. Returning to the example, if we assume conditional independence for the adaptive flexibility construct among adults with autism, this would indicate that the response process within the social adaptive flexibility class is only influenced by their class membership status and nothing else. This may seem like a fairly strong assumption to make; it implies that beyond latent class membership no other factors concurrently explain the covariation between the indicators (e.g., non-modeled variables, additional latent variables, covariates).

Evaluating whether this assumption is realistic can be tested statistically using methods which compare a model with conditional independence assumed to a model with some form of local dependence specified (i.e., residual covariance; Asparouhov & Muthén, 2015). These methods offer a great opportunity for validation testing as evidence supporting conditional independence strengthens the argument that the latent class variable adequately models the joint distribution of the indicators. Nylund-Gibson and Masyn (2016) outline procedures for testing conditional independence in LCA models. Although testing is fairly technical to implement, assessing conditional dependencies in LCA is recommended to ensure that this assumption is justified. When left unevaluated, the conditional independence assumption functions simply as a statistical artifact of the model. However, when evaluated this assumption may provide a more coherent picture of the measurement model's efficacy.

Number of latent classes. A researcher proposing a theory involving a categorical construct must decide how many categories or classes compose the construct to best describe the population under study. In the LCA model, the number of classes is specified by the researcher, and the selected model is assumed to be correctly specified (Nylund et al., 2007). Commonly in applied social science fields, researchers employ an exploratory strategy called enumeration, where a series of models are estimated to determine the number of classes using relative model fit statistics. However, this thesis introduces an alternative approach that utilizes theory-driven conceptual procedures (i.e., confirmatory strategies) to hypothesize the structure of the construct prior to modeling. When taking this approach, enumeration strategies may still be employed; model comparison can be integrated with confirmatory strategies to address validity questions at various stages in the constructing measures process (Schmiege et al., 2017). Presenting a confirmatory approach to LCA aligns with the aim of this thesis by encouraging researchers to confront this central measurement question in the conceptual phase of constructing measures (e.g., What is the structure of the adaptive flexibility construct?).

LCA as a probabilistic model. A fundamental assumption of the LCA model is that it estimates the likelihood (i.e., probability) that observations from the sampled population have membership in each of the nominal classes which compose the categorical construct. This includes the estimation of a distribution for each observation, within which there is a certain probability of correct classification or misclassification across the full set of classes specified. For instance, if the adaptive flexibility construct is assumed to have three classes, the model would estimate the likelihood of each observation belonging to each of these classes (i.e., the posterior probabilities). This suggests that the sample of adults with autism has an imperfect membership in each of the adaptive flexibility classes, indicating that individuals are classified by the model with some degree of uncertainty. In contrast to absolute clustering methods, the LCA model is therefore often described as a probabilistic membership model (Magidson & Vermunt, 2002).

The expression of partial class membership across all specified classes is traditionally explained as meaning that the indicators are imperfect measures of the latent categorical construct or that indicators have some level of measurement error. Alternatively, researchers might consider that the latent categorical construct specified is an imperfect model of the indicator response patterns. Both explanations for measurement error are plausible, as sources of misfit stemming from the latent variable or the indicator responses lead to the same consequence, they are expressed in the posterior probability distribution. In summary, researchers may choose which explanation is most plausible in their research context:

whether the LCA model is incorrectly specified, the indicators are imperfect measures, or both. The probabilistic feature of the LCA model supports a modest commitment about the truth claims of both the models and measures under study, suggesting that classes are neither absolute nor deterministic; Hagenaars and Halman, 1989).

Likelihood of class membership and class size. Once it is assumed that a categorical construct has a finite number of classes a natural question that follows is: What is the relative likelihood of being in one class versus another? This relative class proportion is captured in the LCA model by estimating logit parameters that express the likelihood of being in a focal class relative to the remaining classes (i.e., intercepts). This parameterization aligns with a probabilistic view of the LCA model. However, because absolute class counts provide a convenient interpretation an additional metric of *class size* is commonly reported which is based on sorting observations by most likely class membership (i.e., modal assignment). Modal assignment reflects an absolute classification metric as each observation is sorted into one and only one class from the finite class set. This alternative statistic, although a useful heuristic, provides a contrasting perspective to the probabilistic view of the LCA model and may complicate interpretation as the use of metrics for absolute class size and relative class proportions have different implications.

Although the formal LCA model formulae is unequivocally a probabilistic model, subtle semantic details such as how relative class size is described may lead to misunderstanding in its application. Additionally, a commonly referenced statement in the methods literature is that the LCA model assumes that class membership is both *exclusive and exhaustive* (i.e., each observation is sorted into one and only one class from the specified set). Classes can be considered *exclusive* in the sense that each observation is assumed to have some probability of correct classification in one class as well as some probability of misclassification in the remaining classes. This distinction highlights the difference between the LCA model and the *grade-of-membership* model, which explicitly models partial class membership or degrees of membership across multiple classes (Borsboom et al., 2016). The *exclusive and exhaustive* assumption leads to the interpretation that measurement error is responsible for the expression of misclassification probabilities (i.e., the indicators are imperfect measures). However, as stated previously, another plausible source of error is that the model itself is imperfect, suggesting that the distribution of class probabilities may indicate that the classes are not perfectly differentiated. In this latter case, the LCA model may be viewed as only weakly *exclusive*, as although partial class membership is not explicitly modeled, it is expressed in the posterior probability distribution. This implies that the boundaries between classes may not be perfectly distinguished or that fuzziness exists in the definition of the categorical construct. In this sense, it could be argued that some subset of the sampled population may be characterized by response patterns which do not clearly differentiate between two or more classes. Response patterns which are not clearly differentiated by class bring into question the validity of both models and items.

Notation guide for LCA. This section presents the notation used to describe the LCA model, supplemented by an example to contextualize each syntactic element to its use in practice. To illustrate the formal structure of the LCA model, a simplified version of the adaptive flexibility construct is used to tie syntactic elements listed in Table 3 to a realized context. Consider a categorical latent construct adaptive flexibility. The latent variable is denoted as C_K , where uppercase K represents the total number of classes, and lowercase k specifies a particular class in the finite set. For pedagogical purposes, a simple construct is illustrated, composed of two classes (K=2): a global flexibility class (k=1) and a social flexibility class (k=2). The adaptive flexibility construct is measured by three dichotomous indicators– social, geographic, and topical flexibility (u_1, u_2, u_3) with the response options (m_1, m_2) scored as 0s and 1s indicating either inflexibility or flexibility across each domain. Indicators are denoted by u_i (u_1,..., u_i) with i indicating the total number of indicators constituting the latent variable. For instance, an adult with autism who reports flexibility (m_2) on the social indicator (i.e., $u_1 = 1$) and exhibits the response pattern ($u_1 = 1$, $u_2 = 0$, $u_3 = 0$) would be estimated by the LCA model to have a high posterior probability of having membership in the social flexibility class (k=2).

Table 3

Notation Guide for Latent Class Analysis Model with Dichotomous Indicators

Notation	Description
C_K	Latent class variable
K	Total number of classes
k	Specified class
$u_1,..., u_i$	Indicators of latent class (items)
m_1, m_2	Response options (either 0 or 1)
τ_{ik}	Item and class specific threshold parameters (logit scale)
$\pi_1,..., \pi_k$	Class size probabilities

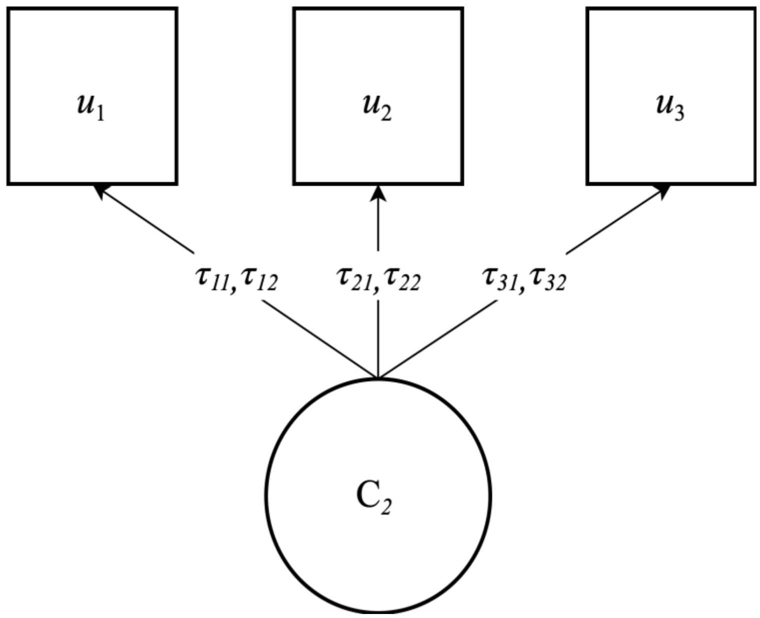
In the LCA model, the relation between a respondent's position on the latent variable (k) and the expected conditional response pattern is described as $pr(u_1,..., u_i|k)$. As the indicators (dependent variables) are dichotomous, this association is modeled using binomial logistic regression paths, which describe how C_K relates to u_1 - u_3. This model form results in the estimation of a set of item- and class-specific logit thresholds denoted as τ_{ik} (i.e., τ_{32} is the threshold for item 3 and class 2). The LCA model form for dichotomous indicators can be represented by the following equation,

$$P(U_1 = m_1, \ldots, U_i = m_i \mid C_K = k) = \prod_{i=1}^{I} \left(\frac{e^{\tau_{ik}}}{1+e^{\tau_{ik}}}\right)^{m_i} \left(\frac{1}{1+e^{\tau_{ik}}}\right)^{1-m_i}.$$

For the adaptive flexibility model with three items and two classes, a total of six threshold parameters (i.e., $\tau_{11}, \ldots, \tau_{32}$; see Figure 3) are needed to describe how indicator responses relate to latent class membership. To estimate relative class size, or the conditional probability that an observation is a member of class k, K-1 logit intercepts are estimated which are then converted to probabilities for ease of interpretation. Class size probabilities are denoted by $\pi_1, \ldots, \pi_k$, so for the adaptive flexibility construct if the social flexibility class was estimated to be 30% ($\pi_1 = .3$) then the global inflexibility class would be the remaining 70% ($\pi_2 = .7$). Parameters which describe the relation between the latent variable and the indicators ($\tau_{11}, \ldots, \tau_{32}$) are conventionally described as *measurement parameters*, while class number (K) and class size proportions ($\pi_1, \ldots, \pi_k$) are distinguished as *structural parameters*.

Figure 3

LCA Path Diagram: Illustrating Parameter Notation for a Two Class Model Composed of Three Dichotomous Indicators



Chapter 2: Defining categorical measures

Proposing research goals.

By clarifying and documenting the motives for measuring a categorical property, the scope and direction of research goals can be refined, leading to a precise construct definition. The choice to define a construct categorically, among alternative construct types, requires justification as this is a fundamental measurement decision (Wilson, 2003). A well-defined research goal not only guides subsequent analytical decisions to justify the measurement of a categorical construct but also forms the foundation for a robust validity argument. In this thesis, an appeal is made for increased emphasis on transparently documenting research goals within applied LCA publication settings. The analysis of the reviewed LCA studies (Appendix B) reveals a frequent under-determination of construct definitions, highlighting a need to strengthen the linkage between research aims and construct definitions. The prevailing use of exploratory methods in applied LCA research (96% of reviewed LCA applications) suggests an opportunity for future studies to integrate specific, testable hypotheses to validate theories. This progression from exploratory to confirmatory research is an important step for any research program; confirmatory studies solidify theories initially proposed through exploratory means. The clarity of research goals and the specificity of construct definitions are essential for drawing valid inferences in categorical measurement, thereby supporting a rigorous application of latent variable theory (Maul, 2017; Messick, 1995).

Adaptive flexibility research goals. The motivation for proposing the adaptive flexibility construct is threefold: 1) To highlight patterns of flexibility and inflexibility across key adaptive behavioral domains; 2) To develop an instrument to measure subtypes or classes of behavioral flexibility in adults with autism; and 3) To challenge the assertion that heterogeneity in autistic populations is sufficiently described by a single spectrum of severity (Constantino et al., 2004). A central aim is to to capture the complex interplay of adaptive and contra-adaptive (restrictive) behaviors manifested across the broad population of adults classified as autistic, highlighting the nuanced diversity within this group. One key benefit of examining class patterns across multiple domains is that it provides a means to deconstruct and describe differences within populations, such as those with autism, moving beyond the simplistic view offered by a singular continuum focused solely on degrees of severity (Happé & Frith, 2021). Exploring categorical constructs offers the opportunity to create new descriptors (i.e., class names) that help identify individual differences in contexts where populations are typically lumped together into a single, uni-dimensional category.

Situating adaptive flexibility in the literature. It is important to situate the adaptive flexibility construct relative to prominent themes established in autism and education research fields. One of three core features defining the autism diagnostic criteria is termed *restricted and repetitive behavior* (RRBs; Leekam et al., 2011) which encompasses behaviors overlapping with those addressed by the adaptive flexibility construct. A brief overview of the extensive literature on RRBs is pertinent to justify the development of this new construct. Additionally, a range of terms are used within autism and education research to describe adaptive and contra-adaptive behavioral observations, including words such as circumscribed, restrictive, and perseverative among others (Juijias et al., 2017). For clarity and consistency, this thesis will use the terms *flexibility* and *inflexibility* to refer to variation across adaptive behavior domains, and readers are encouraged to consult the literature for a broader discussion on this topic (see review; Turner-Brown et al., 2011).

Precedent exists to suggest that the broad RRB classification encompasses qualitatively distinct domains or sub-constructs, as argued in a series of review studies (see Table 4). Research in this area has primarily relied on factor analytic and qualitative methods, advocating for the division of the RRB criteria into various factors such as repetitive sensorimotor behaviors, insistence on sameness, and circumscribed interests among others (Juijias et al., 2017). The adaptive flexibility construct aims to capture key qualitative distinctions across domains of adaptive functioning, emphasizing population heterogeneity in a holistic manner. This shift in theoretical framing and measurement goals, which seeks to identify categorical typologies of adaptive flexibility, necessitates a fundamental redefinition of the construct that conceptually diverges from existing measures. Notably, no single scale reviewed in the RRB literature includes items which provide coverage of all of the following adaptive domains; interest flexibility, social flexibility, geographic flexibility, order flexibility, and sensory flexibility. Therefore, this thesis draws from a wide array of sources in autism and education fields to inform the development of the adaptive flexibility construct.

Table 4

Review Studies of Restrictive and Repetitive Behavior Construct

Review Article	Methods Cited	Proposed Domains/Factors of RRBs
Juijias et al. (2017)	Factor analysis; Qualitative	1) Repetitive sensory motor; 2) Insistence on sameness; 3) Restricted interests
Leekam et al. (2011)	Factor analysis; Qualitative	1) Restricted interests; 2) Nonfunctional routines or rituals; 3) Repetitive motor mannerisms; 4) Preoccupation with parts of objects
Harop et al. (2019)	Single subject design; Qualitative	1) Circumscribed interests; 2) Restrictive interests

The rationale for developing a scale specifically for adult populations with autism stems from the significant underrepresentation of this age group in autism and special education research (Chowdhury et al., 2010). Popular scales for measuring attributes of autism in adults are often adaptations of instruments originally developed to measure pre-adult populations (Bolte, 2012). Adult populations, of sufficient developmental age, are capable of self-reporting, which allows for participant-advocated validity checks. Qualitative methods, such as structured interviews (Maul, 2018), offer adults with autism the opportunity to contribute to the development and validation of measures that describe their behavior. Targeting adult populations may unveil unique patterns of adaptive and restrictive behaviors that are overlooked when using scales designed to measure child or pre-adult populations.

Recent studies affirm the efficacy of self-report measures in adults with autism, demonstrating that this group can accurately report on their behaviors (Huang et al., 2017). Comparisons between self-reported and parent-reported measures in ASD samples show strong correspondence, comparable to those in typically developing populations (Pisula et al., 2017; Schriber et al., 2014). Moreover, attributes associated with behavioral flexibility, such as “insistence on sameness” and “restrictive behaviors”, are noted for their stability and prevalence into adulthood (Chowdhury et al., 2010).

The research questions posed in this applied example aim to explore issues relevant both to the broader categorical measurement goals of this thesis and substantive issues pertinent to understanding heterogeneity in adult populations with autism:

1. Is there empirical support for the five a priori confirmatory classes proposed below in Table 5.
2. Are the identified classes of adaptive behavioral patterns well defined and in agreement with the hypothesized class form– do the confirmatory classes have high within-class homogeneity and between-class separation?
3. Do classes exhibit complex topographies of adaptive and contra-adaptive behaviors simultaneously? If so, can reliable and clear descriptive labels be established to classify these profiles?

4. Are classes of adaptive flexibility significant predictors of functional outcomes and response to intervention?

Table 5

Confirmatory Patterns of the Adaptive Flexibility Construct Classes

	Behavioral domain area				
Class label	Interests	Social	Location	Order	Sensory
Social/interests flexibility	flexible	flexible	inflexible	inflexible	inflexible
Environmental flexibility	inflexible	inflexible	flexible	flexible	flexible
Place/sensory inflexibility	flexible	flexible	inflexible	flexible	inflexible
High flexibility	flexible	flexible	flexible	flexible	flexible
Low flexibility	inflexible	inflexible	inflexible	inflexible	inflexible

Research themes motivating LCA research. The use of LCA methods in applied research is motivated by a variety of common themes and purposes. This thesis includes examples of such purposes, offering researchers guidance in formulating compelling arguments for employing LCA methods. For instance, a common reason for measuring categorical constructs is to provide a description of variation or heterogeneity across sub-groups within a sampled population. In contrast to conventional summary statistics (i.e., means and variances) the LCA approach provides a categorical summary of variation that, by its very nature, produces results that are intuitive and broadly accessible to both academic and non-academic audiences. The colloquial use of terms such as *kinds* and *typologies* to describe population differences underscores the suitability of the LCA model for distilling complex response patterns into a concise and universally comprehensible summary of variance within a population. The importance of research findings being accessible should not be understated; the value of a research product is directly linked to the clarity in which it can be communicated.

Prior to the design of a measure or survey instrument, it is essential to establish a theory-driven definition of the construct. The specificity of this construct definition may vary, depending on the availability of supporting literature, ranging from detailed hypotheses about class structure to more general definitions where literature is limited. This initial conceptual phase of measure construction may begin with informal, ad-hoc efforts to articulate and refine the construct, followed by a more structured process, such as the development of a construct map, as will be elaborated in the following section.

Construct definition and the categorical construct map

Drafting an initial construct definition. After clearly articulating research questions and motives for embarking in the process of constructing a measure, the qualitative work of drafting an initial construct definition can begin. Early definitions of the construct may simply constitute writing a series of notes and ideas to formulate a starting point to seed the construct. Continuing with the ongoing example a preliminary definition for the Adaptive Flexibility construct may look as follows:

The adaptive flexibility construct is defined by a series of behaviors identified as important for adaptive functioning for adults with autism. This construct's intended use is to measure qualitatively distinct classes of individuals based on topographies of behavior that either enable or restrict their ability to flexibly navigate across social and non-social environmental contexts. Five confirmatory class patterns are hypothesized as described in Table 5.

The categorical construct map. The process of defining a construct is elaborated in various measurement texts, often tailored to specific applied contexts such as the development of standardized tests (Mislevy, 2014). These methodologies serve as valuable references for building a toolkit to construct a categorical measure. This qualitative stage of categorical construct definition may be guided by the metaphor of the "construct map" (Wilson, 2005). The construct map is a visual tool used to conceptualize the theoretical space captured by a construct. Originally designed to depict variation in a continuous latent variable, the construct map concept is adapted in this thesis for use with latent categorical variables. A revised version of the construct map, suitable for categorical measures, will be introduced in the subsequent discussion.

The forthcoming section will present the construct map in the form of a series of tables, figures, and diagrams, collectively outlining a strategy for defining a categorical construct. Researchers are encouraged to document each stage of the construct map process showing the conceptual evolution of the construct. To facilitate documentation, it is recommended that researchers create a new annotated copy of construct map tables or figures with each modification. This will provide a clear roadmap detailing the rationale behind each decision made during the construct development process.

Construct domains. At this stage in the construct map process, the sub-components of the construct are designated by the term ‘domain’. Each domain can be thought of as a seed for a potential item, and corresponds to a substantive region of the construct, provided that it is retained through to the final iteration of the construct map. Categorical attributes, such as those distinguishing typologies of individuals within a population, may be considered to be composed by a constellation of sub-attributes (i.e., domains). For example, diagnostic classifications in psychology are characterized by multiple features rather than a single propensity. In this way, categorical constructs can be seen as consisting of interrelated sub-attributes which collectively describe the overarching classes of the construct.

When developing a new construct, candidate domains may vary in their level of association within the domain set. For instance, the adaptive flexibility construct may include interrelated domains targeting aspects of social flexibility and a distinct domain focused on geographic flexibility. Substantively related domains are referred to as a ‘domain area’ in this thesis. A preliminary representation of the adaptive flexibility construct is depicted in Figure 4, where content-overlapping domains are highlighted by shared color-coded cells (e.g., social flexibility in familiar and unfamiliar contexts). The inclusion of relatively distinct domains effectively broadens the scope covered by the construct to a greater degree than a highly associated domain. Theoretically, the inclusion of content-overlapping domains translates to that domain area having greater emphasis or importance in its relative contribution to the construct.

In the review sample of applied LCA studies in Education the average number of indicators was 10 and ranged from 3 to 37 (see Appendix B). Although this may simply reflect an arbitrary discipline-specific convention it likely is related to the fact that larger

LCA models require prohibitively high sample sizes to estimate. In contrast, it is common for continuous latent variable models such as factor analysis or IRT to be applied to instruments of much greater size (e.g., TIMSS; Glynn, 2012). One possible explanation for this phenomenon is that continuous latent variable models often include moderately or highly correlated items with greater overlap in item information contributed by each item (Clark & Bowles, 2015). However, this same statistical trend is not evident in the application of mixture models, high indicator associations are not found to relate to improvements in the models classification quality or model fit (e.g., Entropy, BIC; Bartholomew et al., 2011). Future simulation studies are warranted to evaluate how classification quality and model fit in LCA models relates to indicator association. This comparison is made to emphasize that in the context of categorical constructs the inclusion of associated or content-overlapping domains requires explicit theoretical justification. In conclusion, the decision to include a domain in the construct map, whether distinct or highly associated, should be based on careful theoretical considerations.

Figure 4

Illustrating the 'Domain' and 'Domain Area' Distinction Using the Adaptive Flexibility Example

Domain Area 'Location'	Domain Area 'Social'		Domain Area 'Interest'	
Domain 1	Domain 2	Domain 3	Domain 4	Domain 5

Location flexibility in new environments	Social flexibility in unfamiliar contexts	Social flexibility in familiar contexts	Interest flexibility in conversation	Interest flexibility in activity

In the following section, as we progress through the stages of the construct map, the inclusion or exclusion of domains can occur at any point. Starting with a small set of domains to seed the construct, potential new domain areas may become apparent through engagement with topical literature and the study population. For example, in our review of scales applied to constructs related to adaptive flexibility (e.g., RRBs) it became evident that the interaction between sensory stimuli and the environment was an important area of adaptive behaviors to include for this population.

Implementing the categorical construct map. The construct map process begins by taking the preliminary definition of the construct and systematically refining its structure. Table 6 outlines a multi-step procedure which is intended to guide researchers; the steps, while presented in order, can be adapted or repeated iteratively based on the applied context. Initially, domains are selected intuitively and liberally to sketch a broad, preliminary outline of the construct (steps 1-2). This sets the stage for further refinement where, in subsequent steps, domains are systematically added, evaluated, and possibly removed (steps 2-3). Researchers are encouraged to utilize practitioner experience and observations of the target population during this process.

Table 6

Phases of Categorical Construct Development

Phase	Title	Description	Applied example (Adaptive Flexibility)
Phase 1	Draft Construct Map (concept seed)	List profile patterns based on practitioner observations. Starting with a minimal case is recommended (i.e., 2-3 domains).	Domains: 1) Social; 2) Geographic; 3) Topical interests
Phase 2	Expand Construct	Add domain (columns) to the map. Goal is to make construct coverage comprehensive in-line with theoretical bounds. Each domain can be seen as a candidate item (proceeding the item design phase)	Split social flexibility domain into two: 1) Social flexibility in unfamiliar contexts; 2) Social flexibility in familiar contexts
Phase 3	Refine Construct	Cut redundant or highly overlapping domains. Columns with high content-overlap can be identified by strong correspondence with other domains on the map. Note- items may continue to be refined or cut at a later point.	
Phase 4	Test & Revise Construct	An iterative process of pilot testing, revising, and updating the construct map. Steps 2-3 may require repeating as a cyclical process.	

Exploring class patterns. The method proposed to map out a latent categorical construct is to explore potential class patterns across the topic space. This strategy utilizes qualitative insights, informed by practitioner observation or theoretical precedent, to create a hypothesized variation space (Table 7). For example, in studying adaptive flexibility, a researcher could leverage their clinical experience to identify behavior patterns across key domains such as social preference, social engagement, and interest stability. During this process, theory may be updated or amended to align with each new qualitative observation. Table 7 demonstrates the first step of the construct map process for the adaptive flexibility construct, with columns representing domains and rows indicating observations. Researchers are advised to start with a limited number of construct domains (i.e., 2-3), keeping the response space small initially to facilitate thoughtful construct development.

Table 7

Construct Map I. Depicting Domains and Qualitative Response Patterns

Observation	Social preference	Social engagement	Interest stability
Individual A	Yes	Yes	Yes
Individual B	No	No	Yes
Individual C	Yes	No	Yes
Individual D	Yes	Yes	No

Adding domains. The second step in the construct map process involves adding domains to expand the hypothesized variation space and more comprehensively capture the intended construct. During this stage, construct domains can be added liberally without considering construct size, treating each new domain as a collection of candidate domains. Using an iterative strategy, each new domain added to the construct map table updates the variation space. Researchers are advised to document this development by creating a new table with each modification, thus forming a series of tables that show the evolution of the construct map. This involves duplicating the original table and incorporating additional columns/rows as needed. Table 8 illustrates how adding a 4th construct domain effectively expands the hypothesized variation space. This approach can be seen as a person-centered approach, as domains are added to match individual profiles based on practitioner observations. For instance, a behavioral clinician working with individuals with autism may update the table by adding rows and domains to align with their observations. Following this process, the table may grow wider and longer with each new observation.

Table 8

Construct Map II. Expanding the Construct by Adding Domains

Observation	Social preference	Social engagement	Interest stability	Sensory navigation
Individual A	Yes	Yes	Yes	Yes
Individual B	No	No	Yes	No
Individual C	Yes	No	Yes	Yes
Individual D	Yes	Yes	No	No
Individual E	No	No	No	Yes

Adding domains to a construct increases the potential response patterns exponentially, complicating the prediction of all possible patterns. As shown in Table 3, the exhaustive set of 16 possible response patterns (i.e., 2^4) may not be observed by the practitioner and therefore are not included in the table. Researchers may find that some patterns are more or less plausible theoretically, which will translate, after data collection, to response patterns being either observed or unobserved. To evaluate the theoretical plausibility of response patterns that have not been directly observed in practitioner settings a researcher may add rows for these potential response patterns and consider whether each pattern is plausible on a case-by-case basis. This optional step of the construct map process may be justified by the research practitioner acknowledging that their experience with the population of interest is limited and that a realized sample may reveal patterns not previously encountered. Table 9 below has a set of additional rows (shaded in gray) which are designated as qualitatively "unobserved" indicating that they are potential candidates for inclusion in the construct map response space.

Table 9

Construct Map III. Evaluating the Potential Response Space

Observation	Social preference	Social engagement	Interest stability	Sensory navigation
Individual A	Yes	Yes	Yes	Yes
Individual B	No	No	Yes	No
Individual C	Yes	No	Yes	Yes
Individual D	Yes	Yes	No	No
Individual E	No	No	No	Yes
Individual F	No	Yes	Yes	Yes

Individual G	No	No	No	No
Individual H-P…	*Remaining 9 potential response patterns appended			

This qualitative construct map process may be considered a form of pilot study focused on refining the definition of a construct. An extended version of this process may include the collection of a small pilot sample of qualitative practitioner *observed* response patterns. In this context, following the accumulation of repeated observations, researchers could modify the construct map to include a ‘pattern frequency’ column that records the number of observations per response pattern, as shown in Table 10. This condensed version of the construct map may offer early insights about the relative prevalence of each response pattern. Utilizing this pilot data, researchers can develop early confirmatory hypotheses about the characteristics of latent classes, such as the number of classes, their conceptual distinctions, and their sizes.

Table 10

Construct Map IV: Evaluating Response Frequencies Summarizing Repeated Qualitative Observations

Pattern Frequency	Social preference	Social engagement	Interest stability	Sensory navigation
5	Yes	Yes	Yes	Yes
2	No	No	Yes	No

1	Yes	No	Yes	Yes
3	Yes	Yes	No	No
6	No	No	No	Yes
0	No	Yes	Yes	Yes
0	No	No	No	No

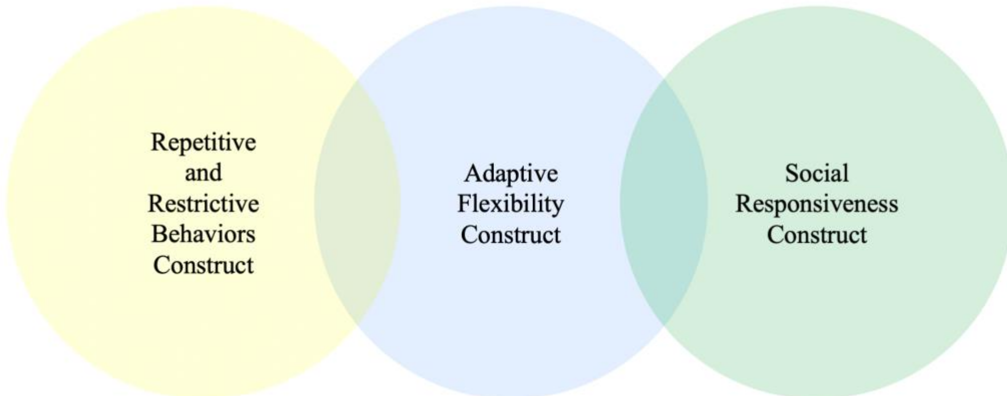
Construct revision. Once a comprehensive set of domains has been collected, construct revision should be considered. Various reasons may motivate the removal of a domain including considerations of parsimony, construct size, and construct boundaries. For instance, a domain may be removed if highly correlated or redundant information is identified across two or more domains (columns). Conversely, a partially redundant or content overlapping domain, may be left in for substantive reasons such as to increase the weight of that domain area's contribution to the overall construct. Such decisions should be made deliberately, based on considerations of theory, and potential redundancies should be identified and evaluated on a case-by-case basis. At this stage, researchers may have a large pool of candidate domains making it pertinent to review and potentially revise construct definitions. This marks a departure from looking at the construct at a granular level to zooming out to evaluate the construct holistically. This shift from detailed analysis to a broader view allows researchers to consider the construct's fit within the wider context of their research area.

Setting construct boundaries. When defining a categorical construct it is crucial to establish its theoretical boundaries, which delineate what is included within or excluded from the construct. The qualitative decision of where to set the constructs boundaries will differentiate the focal construct from neighboring constructs or related but distinct constructs. In this thesis, the term 'construct bounds' is used to refer to the theoretical limits of a construct space. In the educational sciences, social constructs often do not have clearly evident construct bounds (Lamont & Molnár, 2002). For example, the adaptive flexibility construct may be closely related or have overlap with a construct named *repetitive behaviors*. Repetitive behaviors, although not typically considered functional behaviors, have social consequences. Researchers must decide if domains addressing repetitive behaviors should be incorporated into the adaptive flexibility construct or treated as a separate construct. This decision should be based on considerations of the studies research goals and theoretical precedent as definitive or objective bounds may not exist independent of theory.

Figure 5 depicts three related and partly overlapping constructs used to measure behaviors in autism, emphasizing the significance of defining substantive boundaries. Creating a Venn diagram can be a useful exercise for visually assessing the focal construct's boundaries in relation to neighboring constructs, aiding researchers in determining the scope of a categorical construct.

Figure 5

Venn Diagram Illustrating Neighboring and Overlapping Constructs



Construct size. In determining a construct's scope– that is, the number of domains it includes– several factors are considered. Here we introduce the notion of the *level of construct scale* as a pivotal concept that elucidates the role a researcher's perspective plays in shaping and defining the construct. For instance, when defining a construct about bullying a researcher may take an ecologically informed perspective by including domains targeting the greater school and community climate (e.g., Bronfenbrenner, 1994). This *level of construct scale* is distinct from an alternative perspective of the bullying construct that focuses exclusively on acts of bullying at the individual level (e.g. Nylund et al., 2007). Level of construct scale is a closely related but distinct concept from the measurement unit used for data collection. However, the level of construct scale may, in-turn, inform whether bullying data is measured by school-level or individual-level observations. Justification for utilizing a modeling approach such as LCA may be evaluated based on the level of construct scale chosen. A researcher choosing to investigate latent class patterns has made an implicit decision to study a level of construct scale that focuses on group level categories (i.e., classes). These modeling decisions may occlude investigation at lower or higher levels of study (e.g.., individual level, sample level). This modeling tradeoff exposes that complementary perspectives coexist, which may be equally valid, and will highlight alternative facets of the construct. By acknowledging the level of the construct scale, researchers make transparent the role that their perspective plays in defining the construct.

One important consideration when evaluating construct size is the length of the survey instrument required to measure it. The reviewed sample of applied LCA studies featured relatively short instruments compared to instrument sizes seen commonly in continuous latent variable contexts (e.g., Constantino, 2021; Glynn, 2012). While model fit in continuous latent variable models, such as factor analysis, is dependent upon on the strength of inter-item correlations (Clark & Bowles, 2018), in LCA, classification quality and model fit are not driven by indicator associations at the sample level (Maydeu-Olivares et al., 2011).

To illustrate this point, consider a hypothetical dataset consisting of a set of dichotomous indicators with zero inter-item covariances at the sample level– this same data may exhibit clear inter-item dependencies at the latent class level. In contrast, in factor analysis, item correlations directly influence both the loadings within factors and the correlations between factors (Clark & Bowels, 2018). Thus, an LCA model can achieve excellent classification quality and fit with a set of indicators showing minimal associations, which is not the case for factor analysis.

As a result, what qualifies as a *good* item in the LCA framework differs significantly from what is considered a *good* item in factor analysis. This distinction affects item selection: in LCA, a strong solution can be based on items that each contribute a high degree of unique information, whereas in factor analysis, the opposite is true. This statistical distinction may help explain why LCA instruments tend to be shorter than those used in continuous latent variable models. In summary, the LCA model enables researchers to explore complex theories with short surveys, mitigating potential measurement problems associated with longer instruments, such as survey fatigue and response validity issues.

The number of domains and corresponding number of items required to measure the construct should be considered with respect to characteristics of the analytic model. In the context of the LCA model, the number of items included has important statistical and theoretical implications. LCA constructs with few items (i.e., 3 or less) constrain the number of latent classes that can be identified. A general rule is that a minimum of three items is necessary for the latent class model to have stable estimates (Hagenaars & McCutcheon, 2002). Models of this size may adequately describe relatively simple theories which assume the existence of at most two latent classes. A review of applied studies in education and psychology found that LCA models contain an average of 10 items with a range from 3 to 37 items (see Appendix B). Including more items in the model permits the exploration of more detailed and nuanced latent class patterns. Furthermore, expanding the number of items in the construct increases the potential to identify a greater variety of classes, resulting in more intricate analytical outcomes. During the construct map process, the target number of items for the construct should be considered as domains are being added, removed, and assessed.

Making confirmatory hypotheses using the construct map. When following a confirmatory LCA approach, varying degrees of confirmatory hypotheses can be incorporated into the construct map process, contingent on the specificity of theoretical development. Three levels of confirmatory hypotheses, in order of increasing theoretical specificity, are described in Table 11. The gradations include: (1) hypotheses regarding the number of latent classes, (2) hypotheses concerning the shape and size of latent classes, (3) and hypotheses about the meaning and labeling of the latent classes. Researchers may choose the level of confirmatory hypothesis to incorporate based on their theoretical framework. For instance, a researcher might establish hypotheses about the number of latent classes based on prior theory but choose not to hypothesize about the shape or names of the classes before conducting the analysis. This strategy would correspond to a hybrid confirmatory approach, where the shape and naming of classes are determined through exploratory analysis.

Table 11

Three Levels of Confirmatory Hypothesis for Categorical Constructs

	Confirmatory Level	Description
Level 1	Hypothesize number of classes	List the number of class patterns based on established theory or practitioner observations
Level 2	Hypothesize form or shape of classes	Describe the pattern or form of each class, this may include specific hypotheses about expected threshold boundaries
Level 3	Hypothesize meaning and label of latent classes	Describe the meaning of each class and create a descriptive name to label classes

The number of latent classes hypothesized changes the latent variable structurally and may, in turn, influence subsequent confirmatory hypotheses regarding construct shape, size, and meaning. For instance, if a practitioner working with individuals with autism identifies three distinct profiles of social responsiveness, they can tailor the set of domains to align with these typological profiles. In this way, the hypothesized number of latent classes is used to inform decisions regarding the inclusion or exclusion of domains. Furthermore, the shape of latent class patterns may be hypothesized in a confirmatory fashion during the construct map process. To facilitate the definition of confirmatory hypotheses, researchers may draw out a hypothetical class probability plot at this stage of the construct map process (Figure 6). For example, the research practitioner may observe individuals who consistently exhibit behavior that align with a theorized class characterized by endorsement across all domains in the map. A subsequent confirmatory step would be to hypothesize class names to label these groups based on the class patterns hypothesized, such as “high flexibility class”. These proposed class patterns and names must then be validated through data collection and analysis, ensuring they align with theoretical expectations. The process of labeling class patterns is an important substantive consideration and should be adapted or updated in an iterative process following construct development and analysis.

Figure 6

Hypothesized Class Form for Preliminary Adaptive Flexibility Construct

Note. For pedagogical illustration a simplified version of adaptive flexibility construct is presented aligning with the early conceptual phase of construct development.

In instances where the researcher has a priori hypotheses concerning the number and shape of the latent classes an additional step may include revisiting the construct map table (Table 10) to group response patterns according to the anticipated class membership. This entails re-ordering the rows to reflect the expected latent class groupings. Table 12 illustrates a construct map, where rows are arranged by hypothesized class designation, as indicated by the final column and row color. This process of grouping response patterns by expected latent class assignment may prompt researchers to revise hypotheses about the shape or number latent class patterns expected. Entries in the construct map response table directly relate to the hypothesized conditional item probability plot; therefore, any modifications to the expected classes should be reflected in a new iteration of this plot.

Table 12

Construct Map V. Making Preliminary Latent Class Hypotheses

Observation	Social preference	Social engagement	Interest stability	Sensory navigation	**Expected latent class assignment**
Individual A	Yes	Yes	Yes	Yes	Class 1
Individual C	Yes	No	Yes	Yes	Class 1
Individual B	No	No	Yes	No	Class 2
Individual E	No	No	No	Yes	Class 2

Individual D	Yes	Yes	No	No	Class 3

Chapter 3: Item design

The next step in the construct measurement process, item design, takes the conceptual information produced from the construct map process and distills the final set of domains chosen for the categorical construct into survey item questions. This step focuses on developing items that clearly communicate questions to the target population, thus effectively eliciting the intended responses. For instance, an item designed to measure the propensity for social initiation would, if effective, distinguish adults with autism who are flexible in regards to social initiation from those that are inflexible in this social context. This step is a critical piece of the measurement process as survey questions which lead to unintended responses are unlikely to provide meaningful information about the construct and accurately measure the attribute of interest.

Terminology of item design. Before proceeding, it is important to clarify several commonly used terms to avoid semantic ambiguity. In this thesis, the term *item* is used to refer to a survey question with a set of fixed categorical response options. Specifically, the item design process is demonstrated using items with dichotomous response options composed by two categories. While the example of adaptive flexibility uses dichotomous items, the item design procedures discussed can also be applied to measurement contexts that involve more than two response categories, known as multinomial items. The survey item set designated for measuring the construct collectively is referred to as an *instrument*.

Furthermore, item responses are often referred to as *observations*, indicating that the items are the mechanism which elicits an *observation* of the attribute (i.e., it measures). For a general overview of survey theory, including a discussion on various types of item formats, readers are referred to Wilson (2005).

Instruments as context bound. In the case of measuring a categorical attribute, the task of eliciting information that distinguishes observations based on similarities and differences between class typologies cannot be measured directly. Asking participants if they are a member of a 'socially flexible class' directly would be ineffective, as these group classifications are not commonly expressed in our shared lexicon. Instead, we pose questions about specific contexts and the pattern of responses gathered provides a signal indicating class membership. However, it is important to recognize that any given instrument or set of items, designed to capture specific contexts using particular phrasing of questions, does not constitute the exclusive measure of a construct. As Wilson (2005) notes, "The specific questions used are neither necessary for defining the construct nor sufficient to encompass all the possible meanings of the concept [...]" In this way, a construct such as adaptive flexibility is defined uniquely by the set of items chosen from a larger pool of possible items and contexts.

A realist perspective. The empirical aspect of the measurement process thus relies on the presence of a stable attribute—a characteristic of the respondent—that causally influences their response choices. Consequently, the theoretical groundwork which identifies and defines an attribute, which then serves as the target of measurement, provides an anchor point supporting the construct's objectivity. This anchor to an attribute forms the basis for a realist perspective of measurement, the view advocated in this thesis. The realist view of attributes is incompatible with psychological measurement theories such as operationalism and social constructivism (i.e., anti-realist theories). For an in-depth critique of anti-realist measurement theories see Borsboom (2005).

Class form directs item development. When constructing a categorical measure, it is imperative to consider the categorical nature of the attribute during item design. A categorical attribute can be understood as comprising a nominal set of class typologies, with each class describing a pattern of predispositions. Thus, the attribute may consist of a collection of tendencies or interdependent sub-attributes (i.e., domain areas). For instance, individuals may exhibit both flexibility and inflexibility across various social contexts, and the objective of item design is to delineate these patterns of flexibility within the domain space. A well-designed item effectively differentiates individuals based on a context-specific sub-attribute that contributes to each class's respective definition. A class pattern can be seen as a distinctive signature that categorizes individuals into a group with similar combinations of predispositions. A precondition for a class pattern to be identified by the LCA model is that the response pattern must have a sufficient signal or prevalence within the target population. For example, a behavioral pattern observed qualitatively by a researcher may be unique to a specific individual; thus, without repeated observations, it may not have sufficient prevalence to be identified as a latent class in a large quantitative sample. Each item should be assessed for its contribution to shaping the full set of class typology patterns. This necessitates conceptualizing the joint interaction between items that compose the construct collectively. The construct map steps and item design process is oriented to support ways of visually or conceptually representing the joint response space and identify potential class patterns.

Applied example candidate items

For pedagogical purposes, a series of candidate items designed to measure the adaptive flexibility construct are presented here, followed by a detailed discussion of the construction process utilized in designing these items. An initial item pool of 24 fixed response dichotomous items are presented in Table 13 spanning five domain areas– social, interests, location, order, sensory. These items will subsequently be evaluated, refined, and reduced to a smaller subset following an iterative process. As described previously in chapter 2, the purpose for studying this construct is to identify classes of adaptive flexibility in adults with autism to explore patterns of flexibility and inflexibility across key areas identified as functionally adaptive. The intended use for this instrument which aims at identifying categorical class distinctions is to describe the diverse expression of behavioral patterns within this population. This definition of the construct should be considered throughout the item design process. For example, the first item labeled *Primary Interests* is designed to capture inflexibility with regard to a primary interest which occupies a significant portion of the individual's time. In the autism literature, this behavioral tendency is described as a perseverative interest, and is a key characteristic used for identifying autism diagnostically (i.e., DSM-5; American Psychiatric Association, 2013).

Table 13
Survey instructions: For each question, choose if it describes you well or not at all by selecting 'Very much like me' or 'Not at all like me'. Only select the 'Very much like me' option if it really matches how you usually act in the situations described.

Item label	Item wording	Example prompt
Interests		
Primary	I have a favorite interest/activity which I spend most of my time engaged with.	I spend most of my free time making model airplanes.

Change	I recently discovered a new subject or activity that interests me greatly.	I played pickle ball for the first time recently and now I play often.
Interrupt	I would describe myself as inflexible when one of my activities of interest is interrupted or delayed.	When I start working on a new model airplane I will not stop until it is finished.
Initiate	I generally initiate conversations by describing my favorite topic of interest to others.	Hello– have you ever been on a B747 airplane?
Converse	I am flexible when it comes to talking about other people's interests.	I acted interested when my friend brought up baseball, even though I don’t like sports.
Listen	I am a good listener and like to hear about other people’s interests and experiences.	I ask questions about them and wait for their answer.
Social		
Meet	I like to meet new people.	Talking to a person you have met for first time
Groups	I am comfortable in groups of people.	A party or social event
Engage	At a social event I spend most of my time listening or talking to people.	I usually talk to people around me when I’m at a party
Approach	When I see a new person I often introduce myself.	Greeting a person at a social event
Prefer	I prefer to talk to my friends and family.	I would rather talk to my family
Response	I often run out of things to say when the conversation is about other people's interests or experiences.	When it’s my turn to speak I am not sure what to say
Expect	If another person does something unexpected it is easy for me to move past the situation quickly.	A stranger accidentally bumps into you
Location		
New	I like to go to new places that I have never been to before.	I love walking around exploring places I have never been to before.
Same	I prefer to stay in places I am familiar with.	I usually stay at home and always do my shopping in the same store.
Relative	I feel most comfortable when I know exactly where I am relative to places or landmarks that I am familiar with.	I always take the same route to get home.
Order		
Objects	When my things are missing or moved I must always put them back in the proper place.	Every object in my room has a correct place and I rarely move things around.
Sequence	I have routines that I do in a specific order and am uncomfortable if things are done out of order.	I always eat food in a clockwise order

Transition	Transitions from one activity to the next often cause me stress.	I don't like when I have to change activities.
Sensory		
Avoid	I often navigate my environment to avoid specific sensory experiences.	I do not ever walk on grass, instead I will walk around even if it takes a long time.
Control	I am particular about my personal space– choosing surroundings that feel, smell, sound, and look appealing to me.	My room is full of objects that have nice textures.
Seek	When I choose to participate in an activity, sensory details play a big role in my decision making process.	I often choose indoor activities to avoid wearing a jacket.
Touch	I have no problem participating in social conventions that involve touching others.	When greeting others I shake their hand or when it's a friend we hug.

Validation process in item design

In the design of survey items, it is critical that each item is tailored to be as clear as possible to the target population to achieve valid measurement. Direct feedback from this population is the best resource available for refining the language of survey questions. Many sophisticated methods have been developed for gaining insight about the survey response process (Kane, 2006). One such method, *cognitive interviews* (Castillo-Díaz & Padilla, 2013) offers a well tested approach to gather insights about the participants' response processes, aiding in the development and validation of instruments. This method involves conducting structured interviews, and is effective at developing and validating instruments in the drafting phase of item design (Priede & Farrall, 2011). In this thesis, a recently developed method is applied named the *response process evaluation* method (RPE; Wolf et al., 2021) which is less time and resource intensive than structured interview based methods. The RPE method has the advantage of providing a structured approach to document the portion of responses to the probe questions that were interpreted as intended. Furthermore, the RPE method can be used to revise survey items using an iterative procedure as demonstrated in the next section of this thesis. Each of the methods mentioned above utilize direct feedback from the target population to gain insight into the cognitive response process, which is an effective means to integrate input from the participating population into the survey development and validation process.

Incorporating diversified perspectives. Other resources which should be utilized in the item design phase are relevant stakeholders and theoretical precedent. Resources permitting, it is recommended to employ multiple validation strategies as each has its advantages and weaknesses (Kane, 2006). For instance, methods for validation which target the response process may be susceptible to overfitting items to the specific participants sampled. In light of this, items should be checked regularly for alignment with construct definitions and theoretical precedent. Furthermore, items designed to measure categorical constructs should consider the impact each item will have on classification following the structural assumptions imposed when using the LCA model approach.

Expert panel consultation. One such validation strategy employed in the development of the adaptive flexibility construct was to consult with experts in relevant fields to evaluate construct coherence and survey instrument design. A panel of three experts including a psychologist specializing in autism research, a special education researcher, and a measurement methodologist were consulted to review the survey instrument and provide feedback. This feedback was used to inform the next iteration of the adaptive flexibility construct. For example, a discussion with the special education researcher highlighted the importance of framing neurodivergent propensities as differences rather than deficits, prompting a review to ensure that items were sensitive to this issue. Consulting applied researchers, combined with other methods such as response process approaches, demonstrates how diversified perspectives can be incorporated to inform measurement decisions.

The Response Process Evaluation method

To demonstrate how participant feedback can be incorporated into item design the Response Process Evaluation method developed by Wolf et al. (2021) is applied to the ongoing example. This method involves administering structured probe questions designed to elicit respondents' interpretations of survey items. The item "Primary Interests" from the adaptive flexibility construct is presented in Table 14 with respondent feedback provided for illustration. Probe questions aim to elicit insights into how the respondent understands the question, prompting them to provide detailed, open-ended feedback. For instance, in Table 14 the respondent indicated the expected interpretation of the item, providing some support that this item is measuring the target attribute. Assessing the qualitative responses to the *Primary Interests* item collectively, the item interpretation is relatively consistent, responses indicate item comprehension, and there is sufficient response variation. This information provides supporting evidence for the validity of the item suggesting that the item does not require further stages of revision. Such feedback is instrumental for determining if items are interpreted as intended, serving as an effective validation tool for survey design.

Table 14

Questions from the 'Response Process Evaluation' (RPE) method are adapted from Wolf et al. (2023)

How do adults with autism interpret the Adaptive Flexibility survey item *Primary Interests*?
1) Item: 'If these were the response options, which would you select?' *– I have a favorite interest/activity which I spend most of my time engaged with.*'
1) *Response:* "Very much like me"
2) Interpretation: 'In your own words, what does this item mean?'
2) *Response:* "It means I primarily spend my time on the same activity."
3) Example: 'Briefly describe an example of how you typically participate in your favorite interest/activity.'
3) *Response:* "My primary fixations are studying antique porcelain and playing the video game Titanfall."
4) Qualifier interpretation: "In your own words, what does "most of my time engaged with" mean to you?
4) Response: "A larger fraction of my time."
5) Feedback: Is there anything you don't understand or would change about the item? If so, what?
5) *Response:* "No, nothing I can change."

Note. Responses are selected from multiple observations to illustrate informative responses.

Data Collection Procedure for the PRE method. The application of the RPE method involved a multi-stage data collection process. Open-ended data was collected in four stages with survey item format updated and revised at each stage, constituting an iterative instrument development process. The online participant pool *Connect* hosted by Cloud Research (Hartman et al., 2023) was utilized due to its rigorous respondent quality screening process and demographic targeting capabilities. This study, approved by IRB at the University of California, Santa Barbara included adults aged 18-60 diagnosed with autism.

All participants provided consent to participate and were compensated proportionate to survey duration at a rate above the federal minimum wage.

The survey commenced with initial screening questions to verify that diagnoses had been provided by a clinician or medical professional; participants who reported being self-diagnosed were excluded from the qualitative analysis. Feach accompanied by a series of probe questions. Data collection occurred in multiple stages as follows:

1. Preliminary testing: The first stage of data collection consisted of two small test samples (N=5) to ensure survey administration and screening protocols were operating correctly.
2. First collection stage: After this check, the first stage of data was administered (N=20) to investigate ten candidate survey items using a battery of five probes for each item. At this stage items were presented without an example prompt as shown in Table 14 above.
3. Modification (probe adjustment and example prompts): In the next stage of data collection, the same ten items were reassessed with the addition of example prompts and a reduced set of three probes per item (N=20; see Table 15). Modifications were informed by responses from the previous stage with the intention of reducing survey duration, increasing item comprehension, and enhancing the quality of open-ended responses.
4. Final collection phase: The last stage involved presenting nine new candidate items in the same format used in stage 2 but with the reintroduction of the feedback probe used in previous stages ("Is there anything you don't understand or would change about this item? If so, what?"). Additionally, the demographic targeting was adjusted to an age range of 30-60 years, as this strata was found to be less likely to report "self-diagnosed" ASD and returned higher quality responses overall.

Table 15

Revised Survey Presentation Format with Example Prompt and 3 Probe Questions

Revised presentation of RPE method for survey item *Engage Social*
(Probe 1) Item instructions: For each item, choose if the statement describes you well or not at all. Only select the "Very much like me" option if it matches how you usually are in the situations described. Item: "I would describe myself as inflexible when one of my activities of interest is interrupted or delayed." *Example. When I start working on a new model airplane I will not stop until it is finished.* If these were the response options, which would you select? ◯ Very much like me ◯ Not at all like me **(Probe 2)** In your own words, what does this item mean? **(Probe 3)** Briefly describe an example when, "one of your activities of interest was interrupted or delayed".

The adaptive flexibility constructs' purpose (i.e., target of study) was altered in the final phase of data collection to align with participant's with demographic characteristics who returned high quality responses and were readily accessible using the web-based recruitment platform utilized. Namely, the age range for recruitment was shifted higher (from 18-55 years; to 30-60 years) and participants who self-reported an autism diagnosis were excluded. Although this decision to change the target population studied was based on accessibility, given the resources and timeframe feasible for this study, this choice aligns with the primary thesis goal which is to demonstrate a rigorous process to guide in the construction of a categorical measure.

The final sample consisted of 50 participant observations after 8 responses were excluded due to identification of self-diagnosed ASD and low effort response, resulting in an average of 10 observations per-item. This sample size is sufficient for determining whether items are interpreted as expected or warrant further revision. The RPE item development procedure presented using the adaptive flexibility construct provides a template for how the RPE method can be applied in practice. However, as outlined in Wolf et al. (2023), an additional round of RPE testing is warranted to validate the revised items. As this important step of verification was not completed in this study, subsequent research should focus on validating the final version of the instrument to ensure its robustness and applicability.

Analysis and coding coding protocol for RPE data. The RPE response data was analyzed in three phases. In the first phase, qualitative exploratory evaluation, informative responses were highlighted, and key quotes demonstrating the cognitive response process were cataloged. The second phase focused on the fixed-format responses where item variance was evaluated based on response counts ("Very much like me" or "Not at all like me"). Item distributions were used to evaluate whether items effectively classified adaptive flexible predispositions from restrictive inflexible predispositions. Items showing no variance or minimal variance were flagged for potential revision. In the third phase, quantitative coding, a systematic protocol described by Wolf et al. (2023) was applied to evaluate alignment of the responses with the intended construct definition. Responses were coded and categorized as "understood", "not understood", or "not enough information". Responses were considered “understood” if the participants' collective set of responses demonstrated a cognitive response process that aligned with the item's intended meaning. Although coding was solely conducted by the author for this thesis, employing multiple independent coders is recommended to enhance the reliability of the coding results.

Results of RPE data analysis. The results of the RPE coding analysis revealed that the majority of the adaptive flexibility items were understood by participants as intended, based on their responses to the probe questions. For example, the cognitive response process for the item "object organization" presented in Table 16, clearly describes an inflexible predisposition specific to the intended domain. Coherence across probe questions within participants' collective response sets offers evidence that a survey item is effectively targeting the intended measurement construct.

Table 16

Evaluating a Participant Responses Across Probe Questions: Within-Subject Coherence

Item: *When my things are missing or moved I must always put them back in the proper place.*
Probe 1: "To me it means that I prefer organization and that anything out of place creates chaos. If I cannot find the missing item, I will ruminate about it until it is found."
Probe 2: "Once, someone changed the way my globe was facing as a prank. I noticed it immediately and rotated it back to the way I like it."
Probe 3: "The words "must always" imply it must be done immediately and every single time."

In contrast to looking at a participants' collective set of responses, reliability across participants can be assessed by comparing responses for a given probe for each item. Consistency in substantive content across participant responses indicates that the item's meaning was reliably understood within the sampled population. For instance, responses to the probe question for the 'social engagement' item, which asked participants to provide an example of their "favorite interest or activity", consistently aligned with the intended response, fitting the profile of perseverative interests (see Table 17).

Table 17

Probe: "Briefly describe an example of how you typically participate in your favorite interest/activity"

Observation	Response quote
Participant 1	"I spend hours a day reading books and am able to finish several hundred a year. I have a particular interest in Buffy the Vampire Slayer books and if given the choice would read those over everything else."
Participant 2	"I practice my bassoon for long hours and make reeds. I prepare for upcoming concerts. [...]"
Participant 3	"My primary fixations are studying antique porcelain and playing the videogame 'Titanfall'."

Following the scoring method recommended by Wolf et al. (2023), items were scored as meeting a validity threshold if over 80% of responses were coded as "understood". Among the 19 survey items tested, proportions of responses coded as "understood" ranged from 56% to 100%. Overall, 15 items met the validity threshold, with over 80% of responses coded as "understood," while 4 items did not meet this threshold and were subsequently revised.

In summary, the applied example demonstrates how the response process technique can be used to develop and improve instrument quality and provide a disciplined test of item validity. The RPE process outlined above was considered sufficient for demonstrating the efficacy of the method pedagogically. However, further iterations of RPE testing and item development with larger samples are necessary to build a strong evidentiary case for the adaptive flexibilities instrument's validity. In particular, all revised items warrant additional RPE testing to confirm the validity of the instrument in its revised form.

Specificity of Item Language. Item-level semantic specificity is an important consideration when designing items in the LCA modeling context. Categorical latent variables in applied research settings were found to range between 3 and 37 items (see Appendix B). Given this approximate construct size, time may be allocated individually to the design of each item to ensure the measurement target is captured in accordance with theoretical definitions. Furthermore, as indicators in an LCA construct may span a broad area of the construct space (i.e., high inter-item separation), items should be constructed with detailed attention to the intended and unintended contexts (i.e., domains) captured by the construct.

For instance, the item named *Social Engagement* ("At a social event I spend most of my time listening or talking to people.") was written with the expectation that it would encompass contexts of social engagement broadly. Consequently the item wording needs to be general to capture a variety of social engagement styles, yet specific enough to ensure responses have a consistent meaning. Subtle differences in the phrasing of questions may substantially influence how a question is answered. For instance, a respondent indicated that the *Social Engagement* item was "Not at all like me" and provided the following response about whether they understood the item,

> "I am not sure if the listening and talking go together, or could be just one of them. I am mostly silent at events, so the talking part obviously doesn't apply, but I do spend the time listening. [...]"

This response highlights that persons who socially engage primarily by listening (and not talking) may be unsure about how to respond even though listening is intended for this item to constitute a form of social engagement. In light of this response, and low overall variation in response choices across respondents, the item was revised to the following, “At a social event I will <u>typically engage</u> with other people <u>some of the time</u>.”. To further guide respondents on how we expect social engagement to be interpreted a revised example prompt is included with the item, “Although I don’t talk much at social events, I like to listen to what others have to say.”. The example prompt clarifies the intention that listening to others is expected to be interpreted as a form of social engagement. Furthermore, to address the low response variation (9 out of 10 respondents chose ‘Not at all like me’), the qualifier phrase “most of my time” was changed to “some of the time”. This lowers the threshold for item endorsement, with the rationale being that adults with autism are more likely to identify as participating in social engagement “some of the time”, and in consequence will increase item response variation. The semantic details about *what* and *how much* qualifies as being *flexibly adaptive* in the domain of social engagement can significantly alter both the meaning of the responses and the construct.

Measurement specificity is key to providing the detailed information which will differentiate enumerated classes of adaptive flexibility as a clear and distinct set of typologies. The adaptive flexibility construct aims to classify individuals based on their patterns of flexibility across a broad span of behavioral contexts, identifying each context as either adaptive or restrictive. Consequently, each item must be general enough in scope to have relevance to the target population as a whole, while also specific enough to differentiate a nuanced topography of behavioral flexibility. Therefore, each item is designed to target a behavioral context that is pertinent to daily functioning of adults with autism– meaning that when such a behavior is considered restrictive it impacts that person's ability to navigate common social or environmental contexts.

For instance, inflexibility regarding geographic location and navigating novel spaces may restrict prospective employment opportunities. Similarly, geographic inflexibility can limit opportunities for social participation and formation of new relationships. However, geographic inflexibility manifests differently across individuals. The design of context specific survey items is necessary to capture such nuance. For example, two types of geographic inflexibility are expressed in the following responses- one individual expresses a preference for maintaining a familiar reference point at all times, while another states a preference for staying in familiar locations generally (see Table 16).

Table 18

Differentiating types of location flexibility

Item label (interpretation)	Response quote
Relative Location (preference for familiar reference point when navigating a route)	"I need to know where I am at all times and how far away from an exit I am, or at least how far away from a quiet place I am. I also need to know how far away from home I am".
Same Location (preference for familiar location)	"Staying in familiar places is comforting due to an aversion to change."

Careful consideration of the qualifier phrase used in items is important for calibrating the threshold for what constitutes endorsement or non-endorsement of the response choices of a particular item. In the administration of the RPE method, a specific probe-question was included to elicit participants' interpretation of the qualifier phrase in the context of the item (see Table 19). The table below provides an example of how three participants interpreted the qualifier phrase for the "Primary Interests" item and demonstrates why this is consequential for determining the meaning of the response choices. In the case of the item "Primary Interests" the goal is that for participants to identify with the statement as "Very much like me" they must have a strong interest that occupies the majority of their time (i.e., perseverative). The intention is to differentiate individuals whose perseverative interest occupies a significant portion of their time from those who may have a strong interest but effectively balance it with other activities and responsibilities.

Table 19

Calibrating qualifier phrases and meaning of the item response threshold

Probe: In your own words, what does "most of my time engaged with" mean to you? Item: "I have a favorite interest/activity which I spend most of my time engaged with."
Response 1: "Spending an excessive amount of my free time engaged in the activity." Response 2: "A large fraction of my time"
Response 3: "It means what I do 90% of the time I am awake."

A study by Cole et al. (2017) empirically demonstrated the sensitivity of LCA models to variations in item wording. In this study, the authors systematically varied survey item wording and response option format to explore whether the latent class solution would change. The nuanced results of this study, in which the LCA solution changed dynamically by item phrasing and fixed response form, underscore the necessity of making each decision about item presentation thoughtfully and deliberately. This study demonstrates the impact that item design process has on the validity and interpretation of LCA results.

Constructing an item pool. Items can be thought of as being sampled from an "item pool" or "item space" (Wilson, 2004) with the goal of the item set chosen to represent the construct both comprehensively and parsimoniously. In this way, the items designed in any given research context may be considered a finite subset from the exhaustive set of conceivable items (i.e., the universe of possible items; Messick, 1995). Researchers must navigate this item pool, selecting items that adequately cover the intended construct space and adhere to the established construct boundaries. Decisions about the inclusion or exclusion of an item from the item pool will determine the final construct's composition and definition.

In practice, the item pool accessible to a researcher is constrained by practical considerations including economic circumstances, such as time or resources available to dedicate towards the item design process, planned survey format, and the characteristics of the construct under consideration. The item pool is typically restricted by the item format chosen by the researcher, such as items which can be answered with a designated number of fixed response options. For instance, some questions are more suitable for binary yes/no or agree/disagree responses, while others are best suited for a broader range of response options (i.e., multinomial) or an open response format. These considerations must be addressed early in the item design process, as they influence the selection of items for the potential item pool. In this thesis we focus on the simplest case of fixed item response form, namely dichotomous response choices. This puts constraints on the form of question that can be asked to those that have natural correspondence with two categorical response options. In summary, during the item design stage of the construct development process, the item pool for a given research context is constrained by the construct definition, economic considerations, and the response option format determined by the researcher.

Chapter 4: Evaluation and revision of construct

At this stage in the construct measurement process, it is advisable to revisit the initial construct definition formulated prior to item design. This review should assess whether the construct definition requires updating to align with new information discovered in the construct development process, or whether further refinement of items is necessary to fit within the existing theoretical framework. Construct development is a dynamic process; definitions may evolve as researchers engage more deeply with the topic through activities such as the construct map and item design. A comprehensive examination of the instrument is warranted to determine its alignment with the construct definition and stated research goals. This global evaluation is crucial for ensuring the construct's validity and applicability in relation to its intended purpose.

Integral to the adaptive flexibility constructs definition is that each domain area and component sub-domain must clearly distinguish adaptive behavioral predispositions from contra-adaptive ones. Through studying the cognitive response process of the target population via the RPE data, insights were gained about this population's competencies and what constitutes typical behavioral propensities across each of the domain contexts. This information proved invaluable for determining what constitutes a ‘flexible’ or ‘inflexible’ propensity category and how to calibrate survey items to consistently identify such categorical distinctions.

The calibration of the adaptive flexibility instrument based on the participant-informed response data in-turn sharpens the prospective confirmatory latent classes manifest form, fortifying the validity of the categorical construct. In other words, clarifying the meaning and precision of each survey item translates directly to a more precisely defined categorical latent variable, minimizing ambiguity and increasing interpretability of the resultant latent classes. These measurement-oriented principles led to the series of decisions outlined in the next section, detailing the justification for each item revision and instrument modification.

Rationale for Adaptive Flexibility Construct Revisions

A series of revisions were made to the items from the candidate item pool presented previously (Chapter 3; Table 13) based on the RPE results. This section discusses the rationale for each item revision. Some items did not produce the expected interpretation or did not effectively identify categories of flexibility and inflexibility. Thus, these items were either revised or flagged as candidates for removal from the final instrument. The full list of revisions is presented in Table 20.

Table 20

Item Revisions Informed by Response Process Data and Construct Evaluation

Item label *Variation* (*N*)	Item wording for initial item (I) and *revised item (R).*
Engage Soc. *Low (N=10)*	I: "At a social event I spend most of my time listening or talking to people." *R: At a social event I will <u>typically engage</u> with other people <u>some of the time</u>.*
Relative Loc. *None (N=11)*	I: "I feel most comfortable when I know exactly where I am relative to places or landmarks that I am familiar with." *R: "<u>When navigating to a place, I need to know</u> <u>my precise location</u> relative to a landmark or area I am familiar with <u>at all times</u>."*
Avoid Sens. *Low (N=5)*	I: "I often navigate my environment to avoid specific sensory experiences." *R: <u>I regularly go out of my way</u> to navigate my environment to avoid specific sensory experiences.*
Change Int. *Low (N=5)*	I: "I often discover new subjects or activities that interest me greatly." *R: <u>I recently discovered</u> <u>a new interest or hobby</u> that interests me greatly.*
Initiate Int. *None (N=5)*	I: "I generally initiate conversations by describing my favorite topic of interest to others." *R: After introducing myself, I often bring up the topic of my favorite interest when in conversation.*
Response Soc. *None (N=5)*	I: "I often run out of things to say when the conversation is about other people's interests or experiences." *R: I usually run out of things to say when the conversation is about other people's interests or experiences.*
Approach Soc. *None (N=5)*	I: "When I see a new person I often introduce myself." *R: In social settings, I sometimes introduce myself to new people.*
Same Loc. *High (N=5)*	I: "I prefer to stay in places I am familiar with." *R: I prefer to spend time in places I am familiar with.*
Transition Ord. *High (N=5)*	I: "Transitions from one activity to the next often cause me stress." *R: Transitioning from one activity to the next often causes me stress.*

Control Sens. *None (N=5)*	I: "I am particular about my personal space– choosing surroundings that feel, smell, sound, and look appealing to me." *R: Sensory details in my environment often restrict what I do and where I go.*

Note. 'Variation' was classified as follows, *None* = no response variation; *Low* = low response variation; *High* = high response variation. *N* = number of RPE responses.

Following the first wave of RPE data collection and review, all items from the candidate pool were re-evaluated based on the initial results. Three items from the *Social* domain area (*Meet, Group*, and *Prefer*) were removed and did not qualify for testing using the RPE method. This decision was informed in-part by an increased understanding of the target population's social predispositions. For example, the item *Meet Social* was determined likely to produce minimal variance among adults with autism and contribute redundant information due to high substantive overlap with other items (e.g., *Approach*). Additionally, the item *Prefer*, upon re-evaluation, was determined ineffective at differentiating between flexible and inflexible behavior. This is because endorsement of the statement "I prefer to talk to friends and family" is not necessarily indicative of a restrictive behavior, thus the item was flagged as an invalid measure of the adaptive flexibility attribute.

The rationale for making revisions to the *Engagement Social* item has been described in detail in the previous section (see page 67). In summary, this item was not consistently interpreted to communicate that both listening or talking were expected to qualify as social engagement (see Table 20). Additionally, this item had low fixed-response variance, so the qualifier phrase was changed to increase the likelihood of endorsement among adults with autism.

Several items identified through qualitative analysis of the RPE responses did not effectively identify inflexible behavior, particularly behaviors that would be considered contra-adaptive. Items exhibiting this pattern included r*elative location*, *avoid sensory, response social*, *approach social*, and *control sensory*. For example, identifying with the statement “Very much like me” for the item *Relative Location* does not provide a clear signal of restrictive behavior because it is often adaptive to know one’s precise location when navigating (Table 20). To identify individuals who are inflexible due to a heightened sensitivity to their relative orientation, the qualifier phrase was revised to be more specific (e.g., “I need to know my precise location [...] at all times.”). This revision aims to increase the likelihood that participants qualify for non-endorsement of the item (i.e., increase response variation). Items involving relatively idiosyncratic types of behavioral inflexibility, such as *Relative Location*, may be candidates for further revision if they exhibit minimal variance within the larger prospective sample designated for the LCA analysis.

The item *Initiate Social* illustrates an item that was mis-calibrated to the average social awareness of this sample of adults with autism, as the literacy level required to participate in the survey inherently selects for a high functioning population. Although initiations deemed unusual by neurotypical standards have been reported among young adults with autism (Koegel et al., 2014), in contrast, none of the respondents in the RPE sample identified with the initiation style described in the item statement. Participants' cognitive response process for this item collectively reflected a high social awareness that the initiation described in the item prompt would be seen as normatively inappropriate (Table 21). To better align the item with the intended target– identifying individuals who gravitate towards discussing their perseverative interest in conversation– the item was revised. This revision aims to improve the identification of individuals who exhibit inflexibility with regard to conversational subject matter when initiating conversations.

Table 21

Initiate Social item RPE responses illustrates high social awareness within the sample

Observation	Response quote
Participant 1 (probe 1; item interpretation)	"That you inappropriately connect a conversation to an unrelated topic."
Participant 2 (probe 2; item example)	"I'd introduce myself, ask how they are, if it feels appropriate I'd then direct the conversation toward the goal I'm trying to get out of the conversation"
Participant 3 (probe 2; item example)	"If I had to initiate a conversation with a stranger, I would start with sports. A lot of people are also into sports, so it would be a higher likelihood that they might be able to actually have a meaningful conversation with me."

Instrument size and item selection

Instrument size must be evaluated with regard to various conditions specific to the research context including sample size access, the respondent population's survey duration tolerance, and analytic model plan. In the LCA modeling context, instrument size relates proportionately to the number of parameters estimated in the resultant model, making sample size an important consideration. Other considerations include participants' capacity for survey fatigue and the survey engagement propensities of the respondent population.

Furthermore, before conducting quantitative analyses, construct size must be assessed relative to the statistical features of the analytic model chosen. As discussed in previous chapters, LCA models with greater than 20 indicators require sample sizes that may be prohibitively large given the resources typically available in applied social science research (Memon et al., 2020).

Rationale for item selection. Based on the definition of the construct developed previously, the researcher is tasked with selecting items in an effort to represent each domain that comprises the construct in a balanced and intentional way. For instance, when selecting items to represent the adaptive flexibility construct, each of the five domain areas– interests, places, social, order, sensory– should be represented. Choosing a rationale for selecting items and documenting the process is an essential component of a theoretically grounded approach to construct revision.

The process of item selection for the adaptive flexibility instrument is illustrated using the item pool presented in Table 13 from the previous chapter, which initially comprised 23 candidate items. Assessing the final construct size, construct coverage, and sub-construct balance are important considerations. Additionally, items which are expected to have high associations should be identified and evaluated for possible exclusion. A strategy for organizing items based on expected associations is proposed in Table 22, where similar items with overlapping theoretical content are grouped together and their expected association strength and direction is listed. Items with significant substantive overlap may be candidates for exclusion from the instrument. For example, the four items, ‘Converse, Listen, Engage, Approach’, clearly overlap in substantive content and are therefore expected to in-part to redundantly measure the same aspect of the adaptive flexibility attribute. Items under consideration for exclusion based on significant association include, '*Approach, Converse, Interrupt, Same, Response, Seek*' as each of these items theoretically is expected to have a high association with other items in the item pool.

Table 22

Evaluating Candidate Items Based on Hypothesized Association

Item associations	Expected association (Strength / Direction)
Primary, Change	Moderate / Negative
Primary, Initiate	Moderate / Positive
Converse, Listen, Engage, Initiate, Approach	High / Positive
Response, Listen	High / Positive
New, Same	Moderate / Negative
Interrupt, Transition	High / Positive
Objects, Control	Moderate / Positive
Avoid, Seek	Moderate / Negative

To address construct balance between domain areas, a recommended practice is to re-order or group items in a variety of different ways. In Table 23, items are grouped to distinguish between flexibility in social contexts and flexibility in environmental contexts as well as being sorted by domain area. This process of re-sorting items may facilitate decisions to revise or cut items from the item pool to achieve the intended balance for the revised instrument. Items flagged for potential exclusion based on high association are underlined in Table 23.

Table 23

Evaluating Construct Balance: Items Counts Ordered by Social and Non-social Flexibility and Domain Area

Item group	Item count
Social flexibility: <u>*Initiate*</u>, <u>*Converse*</u>, *Listen*, <u>*Meet*</u>, <u>*Groups*</u>, *Engage*, <u>*Approach*</u>, <u>*Prefer*</u>, *Response*, *Expect*, *Touch*	11
Environmental flexibility: *Primary*, *Change*, <u>*Interrupt*</u>, *New*, <u>*Same*</u>, *Relative*, *Objects*, *Sequence*, *Transition*, *Avoid*, *Control*, <u>*Seek*</u>	12
Domain area	
Interests: *Primary*, *Change*, <u>*Interrupt*</u>, <u>*Initiate*</u>, <u>*Converse*</u>, *Listen*	6
Social: *Meet*, *Groups*, *Engage*, <u>*Approach*</u>, *Prefer*, *Response*, *Expect*	7
Location: *New*, <u>*Same*</u>, *Relative*	3
Order: *Objects*, *Sequence*, *Transition*	3
Sensory: *Avoid*, *Control*, <u>*Seek*</u>, *Touch*	4

The qualitative information from the response process evaluation data provided an initial indication about whether theoretical expectations about item associations aligned with realized responses. Although this sample is relatively small, the qualitative observations may provide a validation check either confirming or disconfirming the hypotheses made about item association or redundancy. Through close analysis of the RPE feedback, several items flagged for inter-item associations, after responses were studied, evidence distinct substantive content and therefore are re-considered for their contribution to measuring the adaptive flexibility attribute. For example, participants' interpretation of the item *Interrupt Interests* conveyed that situations when an interest was interrupted provided substantively distinct behavioral flexibility information from the five associated items that comprise the *Interests* domain area. In consideration of the item pair *New Location* and *Same Location,* these items were assumed on theoretical grounds to have a high association after RPE analysis. However, the *New Location* item seems to be interpreted by some respondents as a proclivity to enjoy novel experiences in a general sense. This indicates that the item may be interpreted in multiple ways, potentially altering the meaning of prospective response choice results. The realized interpretation of the *New Location* item when compared with the interpretation of the the *Same Location* item seem to convey different substantive meaning. Consequently, both items were decided to be retained in the revised version of the instrument.

Revised adaptive flexibility construct

Informed by feedback provided by the RPE data, theoretical analysis of sub-construct balance, and overall construct alignment with the construct definition a revised version of the instrument was reached. Such revisions inherently affect the definition of the construct and consequently a revised construct definition is provided below:

> *The adaptive flexibility construct is defined by five domain areas (Social, Interests, Location, Order, Sensory) measured by 17 indicators of which adults with autism are expected to fit in one of five specific behavioral flexibility classes. This construct's intended use is to measure qualitatively distinct classes of individuals based on topographies of behavior that either enable or restrict their ability to flexibly navigate across social and non-social environmental contexts. The class form of the five revised confirmatory class patterns are illustrated in Figure 8.*

A final decision was made to remove six items. The items *Meet*, *Groups*, and *Prefer* were removed prior to RPE analysis as they were deemed ineffective at discriminating flexible behavior from inflexible behavior. The remaining three items *Converse*, *Listen*, *Seek*, were removed based on analysis of item redundancy as informed by the RPE response data. The item counts across domain areas that have social and environmental implications for behavioral flexibility are 8 and 9 respectively for the revised construct (Table 24). The question of whether this difference in items counts across domain areas represents a source of *construct imbalance* is a difficult question to answer without the existence of specific research precedent. Furthermore, upon careful evaluation, each of the nine items targeting environmental flexibility are considered theoretically meaningful to the adaptive functioning of adults with autism. Therefore, at this stage in the revision process, all remaining items are retained.

Table 24

Revised Construct: Item Counts by Domain Area

Domain area	Item label	Item count
Interests	*Primary, Change, Interrupt, Initiate*	4
Social	*Engage, Approach, Response, Expect*	4
Location	*New, Same, Relative*	3
Order	*Objects, Sequence, Transition*	3
Sensory	*Avoid, Control, Touch*	3

The revised instrument, as shown in Table 25, includes 17 items, and is relatively large compared to LCA models found in the social sciences (see Appendix B). However, the author anticipates that post-data analysis the construct can be further revised as needed if items are identified which display limited sample variance or particularly high associations. Furthermore, the theoretical aim of this construct is to highlight behavioral differences across a broad range of adaptive contexts, justifying an item selection approach that prioritizes comprehensive construct coverage.

Table 25

Revised Adaptive Flexibility Instrument Comprised of 17 Indicators

Item label	Item wording	Example prompt
Interests		
Primary	I have a favorite interest/activity which I spend most of my time engaged with.	I spend most of my free time making model airplanes.

Change	I recently discovered a new interest or hobby that interests me greatly.	I played pickle ball for the first time recently and now I play often.
Interrupt	I would describe myself as inflexible when one of my activities of interest is interrupted or delayed.	When I start working on a new model airplane I will not stop until it is finished.
Initiate	After introducing myself, I often bring up the topic of my favorite interest when in conversation.	Hello, my name is Adam. Have you ever been on a B747 airplane?
Social		
Engage	At a social event I will typically engage with other people some of the time.	Although I don't talk much at social events, I like to listen to what others have to say.
Approach	In social settings, I sometimes introduce myself to new people.	At a wedding I made an effort to introduce myself to the bride and groom since they were the hosts.
Response	I usually run out of things to say when the conversation is about other people's interests or experiences.	When it is my turn to speak I am not sure what to say.
Expect	If another person does something unexpected it is easy for me to move past the situation quickly.	A stranger accidentally bumps into you.
Location		
New	I like to go to new places that I have never been to before.	I love walking around exploring places I have never been to before.
Same	I prefer to spend time in places I am familiar with.	I usually stay at home and always do my shopping in the same store.
Relative	When navigating to a place, I need to know my precise location relative to a landmark or area I am familiar with at all times.	I always take the same route to get home and when going to a new location I study the route in detail before leaving.
Order		
Objects	When my things are missing or moved I must always put them back in the proper place.	Every object in my room has a correct place and I rarely move things around.
Sequence	I have routines that I do in a specific order and am uncomfortable if things are done out of order.	I always eat food in a clockwise order
Transitio n	Transitioning from one activity to the next often causes me stress.	I don’t like when I have to change activities.
Sensory		
Avoid	I regularly go out of my way to navigate my environment to avoid specific sensory experiences.	I do not ever walk on grass, instead I will walk around even if it takes a long time.
Control	Sensory details in my environment often restrict what I do and where I go.	I was unable to participate in the birthday party because I am

		sensitive to the smell in my friends house.
Touch	I have no problem participating in social conventions that involve touching others.	When greeting others I shake their hand or when it’s a friend we hug.

Confirmatory research design procedures

To align these construct measure guidelines with a confirmatory approach it is warranted to mention research design and data science practices which align with a confirmatory perspective. Key initiatives aimed at improving the replicability of research findings include the pre-registration movement (Nosek et al., 2021), the open-source movement (Foster & Deardorff, 2017), and study replication efforts (Chhin et al., 2018). These initiatives promote transparency throughout the research process, encompassing data collection, documentation of methodological decisions, and dissemination of results. Such practices are designed to ensure that research studies rigorously test hypotheses while reducing the influence of researcher bias. This body of work serves as an important reminder, particularly relevant to applied LCA research, where exploratory methods are prevalent.

Using an LCA approach for measuring a categorical construct may require a large data collection effort and often involves the sequential estimation of models. Therefore, it is crucial to establish disciplined protocols for the study design plan and to document all decisions made before, during, and after data collection. This section outlines best practices aimed at enhancing research transparency and reducing the potential for researcher bias. Ideally, construct definitions and hypotheses should be explicitly stated and documented before collecting any sample data intended for quantitative analysis. Furthermore, it is advisable to finalize and document the sampling criteria, survey dissemination plan, handling of missing data, and statistical analysis plan before data collection begins. Once these research design details are determined, they should be summarized in a document and, to ensure adherence, can be published on a preprint server (Okon & Ubi, 2021).

Hypothesized class patterns and expected conditional probabilities. Confirmatory LCA analysis entails the formation of specific theory regarding the number of classes and potentially the structure or shape of each latent class. One way to present this hypothesis is to create a depiction of the expected latent classes using the format of a conditional item probability plot. This practice was used during the construct definition phase of the construct map process to formulate an initial confirmatory hypothesis, and included five class patterns (see Table 5). This initial confirmatory hypothesis however was relatively unspecific, including a preliminary theory about how each class is expected to vary by domain area, without specifying item-level detail (i.e., class pattern nuances within domain areas). At this stage, after the revised instrument has been determined, a more specific confirmatory hypothesis is feasible.

Chapter 5: Post-data Categorical Construct Development

This chapter includes an outline of *next steps* in constructing categorical measures following the collection of a quantitative LCA sample. Access to a sample sufficient for descriptive statistical analysis and quantitative modeling (i.e., LCA) presents opportunities for further measurement refinement and construct calibration. Although the collection of a sample large enough for quantitative analysis is beyond the scope of the present work, it is essential to address this topic as a part of a comprehensive guide to categorical measurement within a confirmatory framework.

Data collection considerations. Collecting survey response data of adequate sample size to make persuasive inferences using the LCA modeling approach is not trivial. The resources required to recruit a sample of sufficient size (e.g., N=500; Nylund-Gibson & Choi, 2018) pose a substantial barrier to testing newly developed categorical constructs. This obstacle underscores the importance of conducting rigorous qualitative groundwork to build evidence of the construct's measurement validity before recruiting a quantitative sample.

For a moderately sized LCA construct to obtain adequate power to estimate stable and non-biased parameters, simulation studies have found that 500 to 1000 observations are needed under typical research conditions (Nylund et al., 2007). Factors which contribute to higher sample size requirements include the occurrence of small prevalence classes, indicators with low class discrimination, and escalating complexity of latent variables, such as an increase in the number of indicators and classes.

The review of applied LCA studies demonstrates that applications of LCA often involve more complexity than the simplified conditions simulated in existing literature (see Table 26; Appendix B). For instance, none of the studies cited in Table 26 address the specific conditions present in the Adaptive Flexibility example: five classes, 17 indicators, mixed class separation, and a smallest class size of 6.5%. Note that in Table 26, simulation conditions identified by an asterisk (*) include only well-separated threshold parameters ($|\tau| > .85$); however, applied LCA studies rarely exhibit such ideal class separation across all threshold parameters. This discrepancy underscores the need for future studies to reflect the complexities inherent in applied research settings.

Table 26

Simulated Conditions in Four Monte Carlo Studies of LCA Model Performance

Condition	Yang (2004)	Nylund et al. (2007)	Morovati (2014)	Nylund & Masyn (2016)
Indicators	12, 15, 18	8, 10, 15	7, 10	5
Classes (*K*)	4, 5, 6	3, 4	3, 4	2
Solution complexity	Low	Low*, Moderate*	Ordered, Moderate, High	Low*
Smallest Class	20%	5%	10%	15%

NOTE: Simulation conditions identified by (*) contained only items with high class separation with threshold parameter probabilities >.7 and <.3 ($|\tau| > .85$ logits).

The applied adaptive flexibility example. The adaptive flexibility construct hypothesized exemplifies a complex categorical construct, featuring 17 indicators, 5 classes, and significant variation in item-by-class separation. To demonstrate the calculation of sample size resulting in adequate power specific to the model in question, the confirmatory adaptive flexibility model was tested using a small Monte Carlo simulation study. The simulated data is also utilized as a hypothetical example to demonstrate the steps of a confirmatory LCA analysis and construct refinement process. It is important to note that the power estimates produced using this approach assume perfect model specification. However, given the complexities of real-world data, the assumption that the model precisely replicates the data-generating process is unlikely to hold in applied settings.

Descriptive statistical analysis. After a sample intended for LCA analysis has been collected the observed data can be utilized to further inform construct development. This might take the form of a pilot study or a first study in a series dedicated to refining the construct, time and resources permitting. Construct development which utilizes a data sample and evaluates models to further inform the construct should be differentiated from the conceptual and qualitative processes which proceed quantitative analysis as outlined previously. Examples of descriptive techniques which may be useful in this context include conducting descriptive statistics, examining response patterns, and data visualization methods. Each of these approaches is demonstrated in the following section applied to the adaptive flexibility example.

Using sample data to in-turn inform construct definition has both advantages and disadvantages. A key advantage is that sample data provides opportunities to confirm or disconfirm hypotheses and assumptions, which can then be used to update theory and construct definitions. However, response pattern summaries and preliminary latent class results confirming an existing theory or hypothesis does not prove that the construct is measuring what was intended. Here caution is warranted as data informed decisions are especially prone to inflating our confirmational biases.

A common issue that arises when using data to inform measurement is overfitting (Yarkoni & Westfall, 2017). Overfitting occurs when a model solution is fit to a specific sample's unique characteristics and consequently this solution does not generalize across samples. If a model solution is overfit to a sample, researchers might update the construct definition to fit an eccentricity specific to that sample. In consequence, the overfit construct will no longer have general validity across samples. Keeping these issues in mind, a quantitative sample can provide important insights about LCA constructs and can be a valuable tool to refine theory to correspond with empirical observations.

Examining Response Patterns

The study of response pattern data prior to LCA model specification, although not common practice, may be particularly useful when adhering to a confirmatory approach. Once a sample has been collected, response pattern variation can be effectively visualized using a frequency table format. For instance, Table 27 presents the simulated adaptive flexibility data for the first five indicators displaying response patterns 1-9 (rows appended). In alignment with a confirmatory strategy, the rows in the frequency table patterns can be rearranged to anticipate the composition of latent class groupings, thereby supporting or refining existing hypotheses. An important, though somewhat counterintuitive, aspect of the LCA model is that each prospective latent class is typically composed of multiple response patterns that share overlapping commonalities. Understanding this within-class variation is facilitated by a careful examination of the response frequency table. This exercise serves a similar function to procedures conducted during the construct map phase, where observed patterns are analyzed and hypotheses are either confirmed or amended based on the expectations for confirmatory latent class groupings.

Table 27

Response Frequency Table Ordered by Highest Frequency Pattern from Simulated Adaptive Flexibility Data

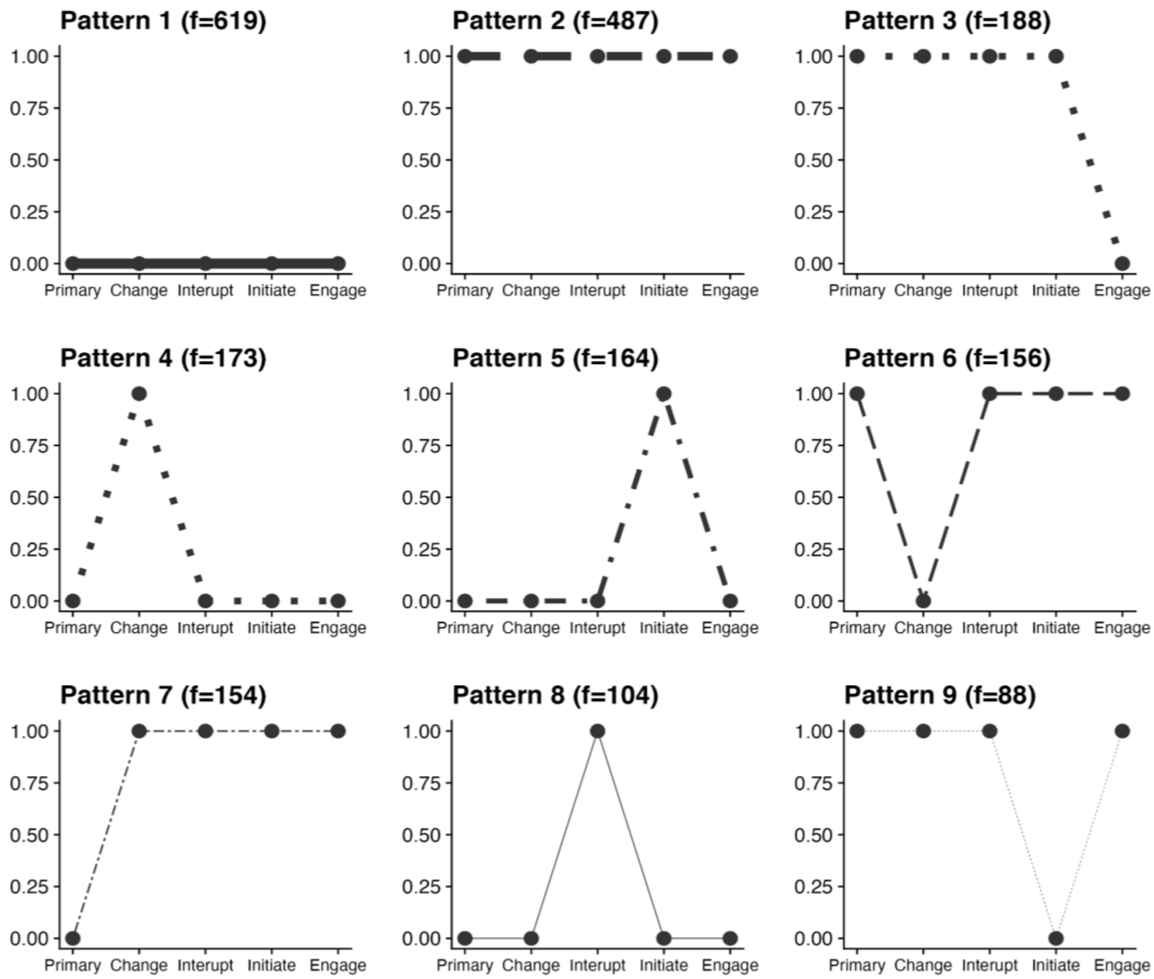
Pattern	Frequency	Item 1	Item 2	Item 3	Item 4	Item 5
00000	619	0	0	0	0	0
11111	487	1	1	1	1	1
…	…	…	…	…	…	…
11101	88	1	1	1	0	1

The response pattern data can be conceptualized as a large contingency matrix in which potential response patterns are either observed or unobserved. In quantitative research, particularly in logistic analysis, unobserved patterns are often referred to as *empty cells* in a large contingency matrix. Due to the exponential increase in potential response patterns with each additional indicator, even a moderately sized construct will typically result in a sparse contingency matrix with a high proportion of empty cells. From a theoretical standpoint, an unobserved pattern can be interpreted in one of two ways: either the pattern was not observed in the sample due to its low prevalence, or the pattern is not exhibited in the population. In some instances, specific combinations of item responses may be contradictory or illogical. The presence of such contradictory patterns in a sample may indicate low-effort responses or data quality issues. These considerations underscore the importance of response pattern analysis in uncovering nuances in categorical response data, establishing it as a valuable step in the confirmatory LCA process.

An alternative method for visually presenting and analyzing response patterns is to use a faceted response pattern plot, as illustrated in Figure 7. This approach, described as a *tile plot*, has the advantage of being parallel in format to the conditional item probability plot, a common means of presenting the results of an LCA solution within social science research. The representation shown in Figure 7 provides a birds eye view of responses which allows researchers to quickly scan across large sets of patterns to identify commonalities and differences across the outcome space. By sorting or reordering facets in the tiled response pattern plot researchers can explore and update theories in an iterative manner. For example, the hypothesized adaptive flexibility class "high social and environmental flexibility" may be composed of 6 unique response patterns characterized by response sets that collectively indicate high adaptive flexibility. These patterns can be arranged according to their association with the hypothesized latent class. Alternatively, tiles can be sorted by frequency to highlight the most prevalent response patterns in the observed sample.

Figure 7

Response Pattern Frequency Tile Plot: Ordered by Highest Frequency from Simulated Adaptive Flexibility Data



Overview of Confirmatory LCA Methods

After exploring the data descriptively the next step is to determine the criteria for specifying a confirmatory latent class analysis (CLCA). As outlined in Schmiege et al. (2017) there are multiple approaches to CLCA which are pertinent across a broad range of theoretical and empirical substantive conditions. In this thesis we will demonstrate both the split sample exploratory-confirmatory method and the hypothesis driven approach to CLCA in-line with the theory-first perspective advocated for in this guide to the measurement of categorical constructs. It is important to emphasize that the example provided is for pedagogical demonstration, given the utilization of simulated data, and that neither of the conditions- existence of sufficient "extant empirical work [and] explicit theoretical advancement" (Schmiege et al., 2017) have been met in the case of the adaptive flexibility construct. Given that these conditions require rigorous theoretical and empirical development specific to the construct form, conditions which are not not met, we will briefly discuss both the hybrid-CLCA approach and hypothesis-driven CLCA approach.

As discussed previously, hypotheses about categorical constructs often fall somewhere along a continuum between exploratory and confirmatory poles. Consequently, hybrid confirmatory/exploratory CLCA analytic approaches may be warranted given the applied research context. For example, researchers may have a hypothesis about the number of classes and their meaning but do not have explicit hypotheses about the degree of within-class homogeneity and between-class separation for each of the constructs' indicators (i.e., boundary hypotheses for τ_{ik}). In this case, a split sample approach can be implemented in which the first sample is enumerated following an exploratory model comparison approach and based on this solution and substantive hypotheses a second sample is analyzed using a confirmatory model with specific parameter restrictions (constraints). Parameter constraints may take a variety of forms, including fixed thresholds, equality constraints, and boundary conditions. This split-sample approach has the advantage of providing empirical confirmation of the hypotheses and then validating the model using the confirmatory sub-sample. For details regarding the determination of constraint specifications using this split-sample method we refer to Schmiege et al. (2017).

The hypothesis driven CLCA approach begins with a confirmatory model in which specific parameter constraints are specified, akin to the analysis of the second sample in the split sample approach. The key difference between these approaches lies in the degree of theoretical development required a priori. This method is particularly suitable when the theoretical foundations are strong and clear hypotheses can be articulated for each parameter. From a measurement perspective that prioritizes theoretical precedent, this approach represents the ideal confirmatory analytic method for validating robust theories of categorical constructs.

Assessing Statistical Power: Simulation Study Results

A statistical power analysis using Monte Carlo simulation methods is a useful tool for determining sample size and as an exercise to understand assumptions of the model and its performance under specific research conditions. Because of this, running a small power simulation analysis is a recommended practice as it helps to ensure that any proceeding models are conducted with sufficient power to minimize estimation bias and produce reliable results.

A replicable example for implementing the small power simulation study found in this thesis can be found in Appendix C. All simulation analyses were conducted in R using the MplusAutomation package with simulated data and models estimated by the Mplus program (R Core Team, 2017, Hallquist & Wiley, 2018, Muthén & Muthén, 2017). The Monte Carlo simulation included nine sample size conditions and 1000 replications were estimated for each condition to ensure stability of the results. A frequently cited guideline in simulation studies suggests that both parameter and standard error bias should not exceed

10% across all estimated model parameters, to ensure robust and reliable results (Muthén & Muthén, 2017). For the hypothesized Adaptive Flexibility construct with 17 binary indicators and 5 classes the estimation of unbiased and stable parameters was reached at a sample size of N=5,000 (see Table 28). These results suggest that larger samples than are currently recommended in the simulation literature (Moravarti, 2014) may be warranted for moderately complex constructs. This preliminary simulation result suggests a need for future simulation studies that cover substantive conditions found in current research applications.

Table 28

Power Study Simulation Results by Sample Size Condition Averaged Across Model Parameters

Condition	Average Bias (%)	Max. Bias (%)	Average SE Bias (%)	Max. SE Bias (%)	Average Coverage	Average Power
N=500	18.80	127.23	32.42	89.13	0.92	0.96
N=1000	5.40	58.39	18.47	80.97	0.95	0.98
N=1500	2.59	29.92	12.10	76.34	0.95	0.99
N=2000	1.46	12.88	7.98	70.83	0.95	1.00
N=2500	1.09	8.41	6.31	64.01	0.95	1.00
N=3000	0.83	4.54	5.34	53.19	0.95	1.00
N=4000	0.63	2.46	3.23	36.86	0.95	1.00
N=5000	0.50	1.70	1.74	5.59	0.95	1.00
N=6000	0.42	1.45	1.94	6.30	0.95	1.00

Note. Average Bias = the percent bias averaged across all model parameters; Max. Bias = the parameter with the largest percent bias across the model parameter set; SE = Standard Error

Confirmatory LCA (CLCA) Implementation and Example Solutions

To demonstrate the split sample method data was simulated with N=6,000 observations providing exploratory and confirmatory samples of sufficient power. All associated code for CLCA analyses are provided in Appendix C. Utilizing the exploratory sample enumeration is conducted for a restricted range of classes (C=3-6) covering the hypothesized class number (C=5) and neighboring class models for comparison. These model results can then be used to compose a table of model fit following normal enumeration procedures- the model fit results for the simulated data are presented in Table 29. Here the model fit results are uniformly in agreement in choosing the 5-class solution, consistent with expectations given the population model specified. The final step in the analysis of the exploratory sample solution is to look at the conditional probability plot and check for correspondence between the estimated conditional item parameters for each class and the hypothesized patterns for the adaptive flexibility construct. Figure 8 presents the five class conditional probability plot for the adaptive flexibility construct.

Table 29

Split Sample Method: Exploratory Sample Model Fit Table for Solutions (K=5, K-1, and K+1)

Model	Par	*LL*	BIC	aBIC	CAIC	AWE	BLRT	VLMR	BF	*cmPk*
4-Class	71	-27,013	54,595	54,369	54,666	55,376	<.001	<.001	0.00	<.001
5-Class	89	-26,202	**53,117**	**52,834**	**53,206**	**54,097**	**<.001**	**<.001**	**>100**	**1.00**
6-Class	107	-26,182	53,222	52,882	53,329	54,400	0.67	0.17	–	<.001

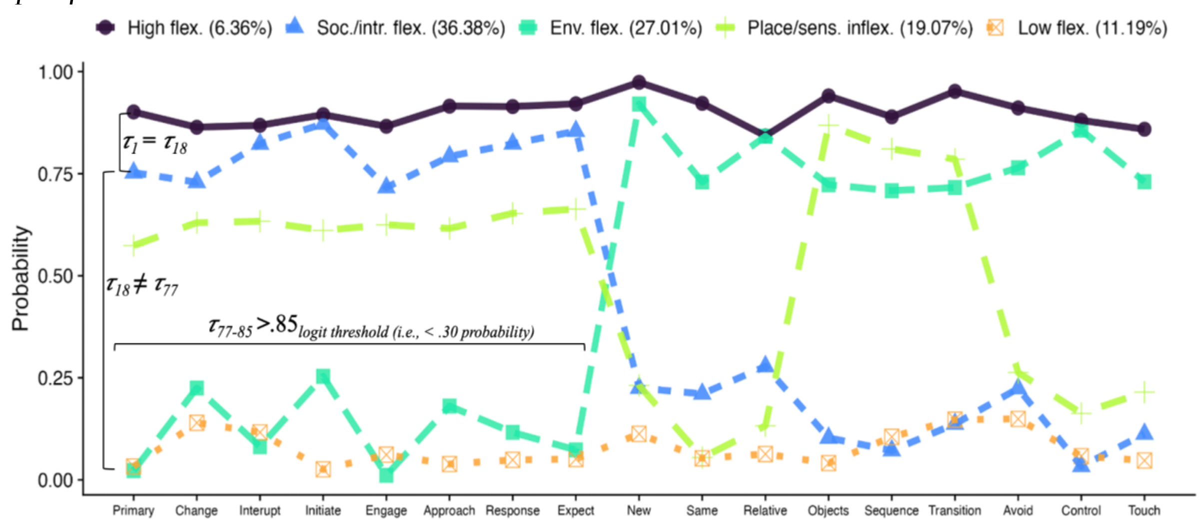
Note. Par = Parameters; *LL* = model log likelihood; BIC = Bayesian information criterion; aBIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion; AWE =

approximate weight of evidence criterion; BLRT = bootstrapped likelihood ratio test p-value; VLMR = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test p-value; *cmPk* = approximate correct model probability

Following the analysis of the exploratory sample, the specification of confirmatory model constraints must be considered in-line with the hypothesized confirmatory model. Using the confirmatory sample, the constraints chosen demonstrate the specification of threshold boundary restrictions for threshold parameters (τ) expected to have clear class separation (i.e., τ <.3 or >.7) and a series of equality constraints for conditional item probabilities hypothesized to be either equivalent or non-equivalent (see; Appendix C). These constraints are illustrated in Figure 8 with selected examples of threshold boundaries, equality and non-equality constraints superimposed on the conditional item probability plot.

Figure 8

Adaptive Flexibility Conditional Item Probability Plot: Illustrating CLCA Constraint Specification



Note. The conditional item probability plot displays the five confirmatory classes of the Adaptive Flexibility construct. The following confirmatory constraints are illustrated: (1) An equality constraint between the thresholds for the Primary item associated with the High Flexibility Class and the Social/Interests Flexibility Class; (2) An inequality constraint between the thresholds for for the Primary item associated with the Social/Interests

Flexibility Class and the Low Flexibility Class; (3) Threshold boundary constraints for the first 8 items in the plot constraining the Low Flexibility parameters to be less than a .30 boundary in the probability scale.

Replication procedures to validate categorical constructs. Replication is a critical tool in the validation of categorical constructs. To test whether a confirmatory sample replicates an initial studies results a multi-group modeling approach can be utilized as outlined in Schmiege et al. (2017). The multigroup procedure is demonstrated using the simulated data for the adaptive flexibility example in Appendix C. This approach allows for both samples to be compared using the estimation of a single model allowing for between group parameter equivalence tests. This procedure provides implementation syntax to conduct a confirmatory replication study in the context of two independent samples. Replication studies are sparse in the applied Social Science literature and it is the hope that the accessible implementation code in Appendix C encourages future replication studies to validate LCA constructs.

Conclusion

The primary objective of this dissertation was to provide a comprehensive guide for constructing categorical measures using a Latent Class Analysis (LCA) modeling approach. This research underscores the importance of developing well-defined measurement instruments to capture categorical properties in social science research. By detailing the iterative steps from construct definition to model validation, this work offers practical guidelines to applied education researchers, particularly those considering the measurement of categorical constructs for the first time. The proposed framework bridges the gap between measurement theory and statistical methods, promoting an integrated approach to construct development that emphasizes validity.

Limitations. Several limitations should be considered when interpreting the findings and recommendations of this dissertation. First, the proposed construct development approach primarily follows a confirmatory strategy, which may not be suitable for all research contexts, particularly those where theoretical frameworks are not well-established. As confirmatory LCA methods require a highly developed theory to specify parameter constraints, their applicability may be limited in exploratory settings where theory is still being formulated.

Second, the dissertation primarily focuses on the measurement of categorical constructs using LCA, without addressing other potential modeling techniques that might be more appropriate depending on the research question or data characteristics. For example, alternative mixture models or grade-of-membership models might offer different insights into population heterogeneity but are not covered in depth here. Future research comparing CLCA and grade-of-membership model approaches would be a useful contribution to the categorical measurement field.

Additionally, the thesis acknowledges the challenge of obtaining sufficient sample sizes for LCA modeling, especially in studies requiring large numbers of observations to achieve stable and reliable parameter estimates. This issue is compounded in contexts where data collection is resource-intensive, potentially limiting the feasibility of the confirmatory strategies outlined.

Finally, the use of hypothetical data to illustrate the CLCA construct development process may limit the generalizability of some of the recommendations provided, as real-world data may present complexities not captured in the examples used. Although this is a significant limitation, the author hopes to have future opportunities to collect a full quantitative sample in the development of the Adaptive Flexibility construct example.

Future Directions. Building on the contributions of this dissertation, several future directions can further advance the study of categorical measurement and the application of LCA models. First, more empirical studies are needed to explore the efficacy of confirmatory LCA methods in various applied research settings. Replicating this study's approach across different contexts, populations, and domains can help validate the guidelines provided and refine the proposed framework.

Second, future research should consider developing hybrid approaches that combine confirmatory and exploratory strategies to accommodate varying contexts of theoretical development. By integrating elements of both confirmatory and exploratory methods, researchers can tailor their approaches to the specific needs and constraints of their studies, enhancing the flexibility and applicability of LCA in practice.

Additionally, further work could explore the application of alternative latent variable models that may better capture the nuances of categorical constructs in certain contexts. For instance, future research might investigate how grade-of-membership models or hierarchical mixture models can be used to assess complex constructs that do not fit neatly within the traditional LCA framework.

Lastly, future studies should emphasize the importance of transparency in documenting research goals, construct definitions, and methodological choices. By providing clear rationales for these decisions, researchers can strengthen the validity of their constructs and enhance the replicability of their findings. Moreover, expanding replication efforts to include diverse populations and settings will provide robust evidence for the generalizability of categorical constructs in social science research.

References

AERA, APA, NCME. (2014). Validity. Standards for Educational and Psychological Testing (p. 11–31). *American Educational Research Association*. https://doi.org/10.1037/e577932014-003

Asparouhov, T., & Muthén, B. (2015). Residual associations in latent class and latent transition analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(2), 169-177.

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Arlington, VA: *American Psychiatric Publishing*.

Balcazar, F. E., Keys, C. B., Kaplan, D. L., & Suarez-Balcazar, Y. (1998). Participatory action research and people with disabilities: Principles and challenges. *Canadian Journal of Rehabilitation*, 12, 105-112.

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). Latent variable models and factor analysis: A unified approach. *John Wiley & Sons*.

Bölte, S. (2012). Brief report: the social responsiveness scale for adults (SRS-A): initial results in a German cohort. *Journal of Autism and Developmental Disorders*, 42, 1998-1999.

Borsboom, D. (2005). Measuring the mind: Conceptual issues in contemporary psychometrics. Cambridge University Press. https://doi.org/10.1111/j.1600-0447.2006.00820.x

Borsboom, D. (2008) Latent Variable Theory, Measurement. *Interdisciplinary Research and Perspectives*, 6:1-2, 25-53, https://doi.org/10.1080/15366360802035497

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203.

Borsboom, D., Rhemtulla, M., Cramer, A. O., van der Maas, H. L., Scheffer, M., & Dolan, C. V. (2016). Kinds versus continua: A review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychological Medicine*, 46(8), 1567-1579.

Bronfenbrenner, U. (1994). Ecological models of human development. *International Encyclopedia of Education*, 3(2), 37-43.

Castillo-Díaz, M., & Padilla, J. L. (2013). How cognitive interviewing can provide validity evidence of the response processes to scale items. *Social Indicators Research*, 114, 963-975.

Chowdhury, M., Benson, B. A., & Hillier, A. (2010). Changes in restricted repetitive behaviors with age: A study of high-functioning adults with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 4(2), 210-216.

Chhin, C. S., Taylor, K. A., & Wei, W. S. (2018). Supporting a culture of replication: An examination of education and special education research grants funded by the Institute of Education Sciences. *Educational Researcher*, 47(9), 594-605.

Clark, D. A., & Bowles, R. P. (2018). Model fit and item factor analysis: Overfactoring, underfactoring, and a program to guide interpretation. *Multivariate Behavioral Research*, 53(4), 544-558.

Cole, V. T., Bauer, D. J., Hussong, A. M., & Giordano, M. L. (2017). An empirical assessment of the sensitivity of mixture models to changes in measurement. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(2), 159-179. https://doi.org/10.1080/10705511.2016.1257354

Constantino, J. N. (2021). Social responsiveness scale. *In Encyclopedia of Autism Spectrum Disorders* (pp. 4457-4467). Springer International Publishing.

Constantino, J. N., Gruber, C. P., Davis, S., Hayes, S., Passanante, N., & Przybeck, T. (2004). The factor structure of autistic traits. *Journal of Child Psychology and Psychiatry*, 45(4), 719-726.

Didden, R., Sigafoos, J., Green, V. A., Korzilius, H. P. L. M., Mouws, C. F. A. M., Lancioni, G. E., ... & Curfs, L. M. (2008). Behavioral flexibility in individuals with Angelman syndrome, Down syndrome, non-specific intellectual disability and autism spectrum disorder. *Journal of Intellectual Disability Research*, 52(6), 503-509. https://doi.org/10.1111/j.1365-2788.2008.01055.x

Duff, M., & Bowers, A. J. (2022). Identifying a typology of New York City schools through teacher perceptions of organizational capacity: A latent class analysis. Leadership and Policy in Schools, 21(4), 791-815. https://doi.org/10.1080/15700763.2020.1854789

Finch, W. H., & Bronk, K. C. (2011). Conducting confirmatory latent class analysis using M plus. Structural Equation Modeling, 18(1), 132-151.

Fink, M. C., Radkowitsch, A., Bauer, E., Sailer, M., Kiesewetter, J., Schmidmaier, R., ... & Fischer, M. R. (2021). Simulation research and design: a dual-level framework for multi-project research programs. Educational Technology Research and Development, 69, 809-841.

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. Social Psychological and Personality Science, 8(4), 370-378. https://doi.org/10.1177/1948550617693063

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF) as a tool for open science practices. Advances in Methods and Practices in Psychological Science, 1(2), 242-253. https://doi.org/10.1177/2515245917747646

Gao, Y., Zhai, X., Andersson, B., Zeng, P., & Xin, T. (2020). Developing a learning progression of buoyancy to model conceptual change: A latent class and rule space model analysis. Research in Science Education, 50, 1369-1388. https://doi.org/10.1007/s11165-018-9736-5

Gerber, M., Wittekind, A., Grote, G., & Staffelbach, B. (2009). Exploring types of career orientation: A latent class analysis approach. Journal of Vocational Behavior, 75(3), 303-318.

Glynn, S. M. (2012). International assessment: A Rasch model and teachers' evaluation of TIMSS science achievement items. Journal of Research in Science Teaching, 49(10), 1321-1344.

Green, L. W., & Glasgow, R. E. (2006). Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology.

Evaluation & the health professions, 29(1), 126-153. https://doi.org/10.1177/0163278705284445

Hagenaars, J. A., & Halman, L. C. (1989). Searching for ideal types: The potentialities of latent class analysis. European Sociological Review, 5(1), 81-96.

Hagenaars, J. A., & McCutcheon, A. L. (2002). Applied latent class analysis. Cambridge University Press.

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: an R package for facilitating large-scale latent variable analyses in Mplus. Structural equation modeling: a multidisciplinary journal, 25(4), 621-638.

Happé, F., & Frith, U. (2021). Dimensional or categorical approaches to autism? Both are needed. A reply to Nick Chown and Julia Leatherland. Journal of Autism and Developmental Disorders, 51, 752-753.

Hartman, R., Moss, A. J., Jaffe, S. N., Rosenzweig, C., Litman, L., & Robinson, J. (2023). Introducing connect by CloudResearch: Advancing online participant recruitment in the digital age.

Huang, A. X., Hughes, T. L., Sutton, L. R., Lawrence, M., Chen, X., Ji, Z., & Zeleke, W. (2017). Understanding the self in individuals with autism spectrum disorders (ASD): A review of literature. Frontiers in psychology, 8, 1422.

Jansen, B. R., & van der Maas, H. L. (1997). Statistical test of the rule assessment methodology by latent class analysis. Developmental Review, 17(3), 321-357. https://doi.org/10.1006/drev.1997.0437

Jiujias, M., Kelley, E., & Hall, L. (2017). Restricted, repetitive behaviors in autism spectrum disorder and obsessive–compulsive disorder: A comparative review. Child Psychiatry & Human Development, 48, 944-959.

Kane, M. (2006). Content-related validity evidence in test development. Handbook of test development, 1, 131-153.

Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. Neuropsychology review, 18, 320-338. https://doi.org/10.1007/s11065-008-9077-7

Kim, D. H., Lambert, R. G., & Burts, D. C. (2018). Are young dual language learners homogeneous? Identifying subgroups using latent class analysis. The Journal of Educational Research, 111(1), 43-57. https://doi.org/10.1080/00220671.2016.1190912

Koegel, R. L., Bradshaw, J. L., Ashbaugh, K., & Koegel, L. K. (2014). Improving question-asking initiations in young children with autism using pivotal response treatment. Journal of autism and developmental disorders, 44, 816-827.

Laudy, O., Zoccolillo, M., Baillargeon, R. H., Boom, J., Tremblay, R. E., & Hoijtink, H. (2005). Applications of confirmatory latent class analysis in developmental psychology. European Journal of Developmental Psychology, 2(1), 1-15.

Lamont, M., & Molnár, V. (2002). The study of boundaries in the social sciences. Annual review of sociology, 28(1), 167-195.

Leekam, S. R., Prior, M. R., & Uljarevic, M. (2011). Restricted and repetitive behaviors in autism spectrum disorders: a review of research in the last decade. Psychological bulletin, 137(4), 562.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. Psychological bulletin, 114(1), 185.

MacCallum, R. C. (2003). 2001 presidential address: Working with imperfect models. Multivariate behavioral research, 38(1), 113-139.

Maydeu-Olivares, A., Cai, L., & Hernández, A. (2011). Comparing the fit of item response theory and factor analysis models. Structural Equation Modeling: A Multidisciplinary Journal, 18(3), 333-356.

Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: A comparison with K-means. Canadian journal of marketing research, 20(1), 36-43.

Markus, K. A., & Borsboom, D. (2013). Frontiers of test validity theory: Measurement, causation, and meaning. Routledge.

Mari, L., Maul, A., Irribarra, D. T., & Wilson, M. (2017). Quantities, quantification, and the necessary and sufficient conditions for measurement. Measurement, 100, 115-121. https://doi.org/10.1016/j.measurement.2016.12.050

Mari, L., Wilson, M., & Maul, A. (2021). Measurement across the Sciences. Springer Ser. Meas. Science and Technology. https://doi.org/10.1007/978-3-031-22448-5

Maul, A. (2017). Rethinking traditional methods of survey validation. Measurement: Interdisciplinary research and perspectives, 15(2), 51-69. https://doi.org/10.1080/15366367.2017.1348108

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling.

Maul, A. (2017). Rethinking traditional methods of survey validation. Measurement: Interdisciplinary research and perspectives, 15(2), 51-69.

Maul, A. (2018). Validity. The SAGE encyclopedia of educational research, measurement, and evaluation, 1771-1775. https://doi.org/10.4135/9781506326139.n731

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American psychologist, 50(9), 741.

Memon, M. A., Ting, H., Cheah, J. H., Thurasamy, R., Chuah, F., & Cham, T. H. (2020). Sample size for survey research: Review and recommendations. Journal of Applied Structural Equation Modeling, 4(2), 1-20.

Muthén, B. (2001). Latent variable mixture modeling. New developmental and techniques in structural equation modeling/Lawrence Erlbaum Associates, Inc.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. Structural equation modeling, 9(4), 599-620.

Muthén, B., & Muthén, L. (2017). Mplus. Handbook of item response theory (pp. 507-518). Chapman and Hall/CRC.

Michell, J. (2012). Alfred Binet and the concept of heterogeneous orders. Frontiers in psychology, 3, 261. https://doi.org/10.3389/fpsyg.2012.00261

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. ETS Research Report Series, 2003(1), i-29.

Morovati, D. (2014). The intersection of sample size, number of indicators, and class enumeration in LCA: A Monte Carlo study (Doctoral dissertation, UC Santa Barbara).

Nasiopoulou, P., Williams, P., Sheridan, S., & Hansen, K. Y. (2017). Exploring preschool teachers' professional profiles in Swedish preschool: A latent class analysis. Early Child Development and Care. https://doi.org/10.1080/03004430.2017.1375482

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural equation modeling: A multidisciplinary Journal, 14(4), 535-569.

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. Translational Issues in Psychological Science, 4(4), 440.

Nylund-Gibson, K., & Masyn, K. E. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. Structural Equation Modeling: A Multidisciplinary Journal, 23(6), 782-797.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. Perspectives on Psychological Science, 13(6), 615-621. https://doi.org/10.1177/1745691618794188

Okon, A. E., & Ubi, I. O. (2021). Awareness and Utilisation of PsyArXiv for the Dissemination of Research Preprints Among Psychology Scholars in Universities. Library Philosophy and Practice, 5870.

Pisula, E., Pudło, M., Słowińska, M., Kawa, R., Strząska, M., Banasiak, A., & Wolańczyk, T. (2017). Behavioral and emotional problems in high-functioning girls and boys with autism spectrum disorders: Parents' reports and adolescents' self-reports. Autism, 21(6), 738-748.

Priede, C., & Farrall, S. (2011). Comparing results from different styles of cognitive interviewing: ‘verbal probing’ vs.‘thinking aloud’. International Journal of Social Research Methodology, 14(4), 271-287.

R Core Team (2017) R: A Language and Environment for Statistical Computing. https://www.R-project.org/

Schmiege, S. J., Masyn, K. E., & Bryan, A. D. (2018). Confirmatory latent class analysis: Illustrations of empirically driven and theoretically driven model constraints. Organizational Research Methods, 21(4), 983-1001. https://doi.org/10.1177/1094428117747689

Schriber, R. A., Robins, R. W., & Solomon, M. (2014). Personality and self-insight in individuals with autism spectrum disorder. Journal of personality and social psychology, 106(1), 112.

Slaney, K. (2017). Validating psychological constructs: Historical, philosophical, and practical dimensions. Springer.

Slaney, K. L., & Racine, T. P. (2013). What’s in a name? Psychology’s ever evasive construct. New Ideas in Psychology, 31(1), 4-12. https://doi.org/10.1016/j.newideapsych.2011.02.003

Wilson, M. (2003). On choosing a model for measuring. Methods of Psychological Research, 8(3), 1-22.

Wilson, M. (2005). Constructing measures: An item response modeling approach. Taylor & Francis. https://doi.org/10.4324/9781410611697-6

Wolf, M. G., Ihm, E., Maul, A., & Taves, A. (2021). Survey item validation. In The Routledge handbook of research methods in the study of religion (pp. 612-624). Routledge.

Wolf, M. G., Ihm, E., Maul, A., & Taves, A. (2023). The response process evaluation method. Preprint. PsyArXiv. https://doi. org/rbd2x

Yang, C. C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. Computational statistics & data analysis, 50(4), 1090-1104.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science, 12(6), 1100-1122. https://doi.org/10.1177/1745691617693393

Appendix A

#	Authors	Article Title	Exist-ing Scale	Seco-ndary Data	Reco-ded Scale	Expl-orato ry
1	Burgos-Videla et al.	Digital Competence Analysis of University Students Using Latent Classes	Y	Y	N	Y
2	Denson, Ing	Latent Class Analysis in Higher Education: An Illustrative Example of Pluralistic Orientation	Y	Y	Y	Y
3	Liu et al.	Application of latent class analysis in assessing the competency of physicians in China	Y	N	Y	Y
4	Choi, Kang	Korean engineering majors' perspective toward lifelong learning using latent class analysis	Y	N	Y	Y

5	Lingo, Chen	Righteous, Reveler, Achiever, Bored: A Latent Class Analysis of First-Year Student Involvement	Y	Y	N	Y
6	Kim et al.	Are young dual language learners homogeneous? Identifying subgroups using latent class analysis	Y	N	Y	Y
7	Nasiopoulou et al.	Exploring preschool teachers' professional profiles in Swedish preschool: a latent class analysis	Y	Y	Y	Y
8	Custer, Akaeze	A Typology of State Financial Aid Grant Programs Using Latent Class Analysis	N	N	N	Y
9	Marraccini et al.	Instructor and peer bullying in college students: Distinct typologies based on Latent Class Analysis	Y	N	Y	Y
10	Denson et al.	A Latent Class Analysis of Students' Openness to Learning From Diverse Others	Y	Y	Y	Y
11	Campbell et al.	From Comprehensive to Singular: A Latent Class Analysis of College Teaching Practices	N	N	N	Y
12	Akçakin, Kaya	Determining High School Students' Mathematical Thinking Styles: Latent Class Analysis	Y	N	N	Y
13	Berkowitz et al.	A Latent Class Analysis of Victimization Among Middle and High School Students in California	Y	Y	N	Y
14	Timmons, Pelletier	Using latent-class analysis to examine the influence of kindergarten children's perspectives of school on literacy and self-regulation outcomes	Y	Y	N	Y
15	Lounek, Ryska	Juxtaposing Acquired and Required Skills: Latent Class Analysis of Self-Assessment Scales in an International Survey	Y	Y	Y	Y
16	Patte et al.	Does school connectedness differ by student ethnicity? A latent class analysis among Canadian youth	Y	Y	Y	Y
17	Kang	Heterogeneity of Learners' Behavioral Patterns of Watching Videos and Completing Assessments in Massive Open Online Courses (MOOCs): A Latent Class Analysis	N	N	N	Y

18	Khan, Krell	Patterns of Scientific Reasoning Skills among Pre-Service Science Teachers: A Latent Class Analysis	Y	N	Y	Y
19	Cai, Cheung	Classifying the writing assessment tasks of English as the medium of instruction programs using latent class analysis	N	N	N	Y
20	Basaran, Yalman	Determining the perceptions of pre-service teachers on technology-based learning during the Covid-19 process: a latent class analysis approach	N	N	N	Y
21	Urick	The influence of typologies of school leaders on teacher retention A multilevel latent class analysis	Y	Y	Y	Y
22	Toker, Green	A Comparison of Latent Class Analysis and the Mixture Rasch Model Using 8th Grade Mathematics Data [...]	Y	Y	Y	Y
23	Stein, Jimerso n	An Examination of Bullying Roles and Moral Disengagement Using Latent Analysis	Y	N	Y	Y
24	Gao et al.	Developing a Learning Progression of Buoyancy to Model Conceptual Change: A Latent Class and Rule Space Model Analysis	N	N	N	N
25	Kang, Choi	A Study on Finding out Barriers of Diffusion of Social Media - Assisted Learning: Focusing on the perception of learners Using Latent Class Analysis	N	N	Y	Y
26	Greiff et al.	Students' exploration strategies in computer-simulated complex problem environments: A latent class approach	Y	Y	N	Y
27	Brown	Using Latent Class Analysis to Set Academic Performance Standards	Y	Y	Y	Y
28	Cho	Bullying Victimization, Negative Emotionality, and Suicidal Ideation in Korean Youth: Assessing Latent Class Analysis Using the Manual 3-Step Approach	Y	Y	Y	Y
29	Agasisti et al.	School principals' leadership types and student achievement in the Italian context: Empirical results from a three-step latent class analysis	Y	Y	Y	Y
30	Karakoy un, Basaran	Identifying Turkish students' profiles of using information and communication technologies and its relationship with their	Y	Y	Y	Y

		academic achievement: A latent class analysis approach				
31	Urick, Bowers	What Are the Different Types of Principals Across the United States? A Latent Class Analysis of Principal Perception of Leadership	Y	Y	Y	Y
32	Jones, Rosenberg	Characterizing whole class discussions about data and statistics with conversation profile analysis	N	N	N	Y
33	Burns et al.	High school students' out-of-school science participation: A latent class analysis and unique associations with science aspirations and achievement	Y	N	Y	Y
34	Zhong et al.	Bullying and Victimization in Chinese Affordable Kindergartens: A Latent Profile Analysis	Y	Y	N	Y
35	Weerts et al.	Uncovering Categories of Civically Engaged College Students: A Latent Class Analysis	Y	Y	Y	Y
36	Fematt et al.	Identifying Transfer Student Subgroups by Academic and Social Adjustment: A Latent Class Analysis	N	N	Y	Y
37	Murtafi'ah et al.	A Latent Profile Analysis of Santri's Reading Attitude and Reading Motivation	N	Y	Y	Y
38	Olivera-Aguilar et al.	Using Latent Profile Analysis to Identify Noncognitive Skill Profiles Among College Students	Y	Y	N	Y
39	Postigo et al.	Academic grit modulates school performance evolution over time: A latent transition analysis	Y	Y	Y	Y
40	Brandriet, et al.	Evaluating students' abilities to construct mathematical models from data using latent class analysis	Y	N	N	Y
41	King et al.	Determinants of Black families' access to a community-based STEM program: A latent class analysis	Y	Y	Y	Y
42	Fleary	Combined Patterns of Risk for Problem and Obesogenic Behaviors in Adolescents: A Latent Class Analysis Approach	Y	Y	Y	Y

43	Bowers, Sprott	Why Tenth Graders Fail to Finish High School: A Dropout Typology Latent Class Analysis	Y	Y	Y	Y
44	Coombs et al.	A person-centered analysis of teacher candidates' approaches to assessment	Y	N	Y	Y
45	Choi, Kang	A Dynamic Examination Of Motives For Using Social Media And Social Media Usage Among Undergraduate Students: A Latent Class Analysis	N	N	N	Y
46	Boutin-Martine et al.	Exploring Resilience in Latina/o Academic Outcomes: A Latent Class Approach	Y	Y	Y	Y
47	Duff, Bowers	Identifying a Typology of New York City Schools Through Teacher Perceptions of Organizational Capacity: A Latent Class Analysis	Y	Y	Y	Y
48	Dang, Nylund-Gibson	Connecting Math Attitudes With STEM Career Attainment: A Latent Class Analysis Approach	Y	Y	Y	Y
49	Harshman, Yeziers ki	Characterizing high school chemistry teachers' use of assessment data via latent class analysis	N	N	N	Y
50	Vaval et al.	Identifying a typology of high schools based on their orientation toward STEM: A latent class analysis of HSLS:09	Y	Y	N	Y
51	Harlow et al.	Using Latent Class Analysis to Analyze Children's Responses to the Question, What Is a Day?	N	N	N	Y
52	Kim, D et al.	Latent class analysis of non-formal learners' self-directed learning patterns in open educational resource repositories	Y	Y	N	Y
53	Zhu, YT et al.	Exploring Patterns of Self-control and the Relationship with Home-rearing Environment Among Preschoolers	Y	Y	Y	Y
54	Weissinger, et al.	Barriers to mental health services among college students screened in student health: A latent class analysis	Y	Y	Y	Y
55	Helsabeck et al.	Pathways to Kindergarten: A Latent Class Analysis of Children's Time in Early Education and Care	N	N	N	N

56	Gutiérrez et al.	Early Prediction of Reading Risk in Fourth Grade: A Combined Latent Class Analysis and Classification Tree Approach	Y	Y	Y	Y
57	Poesen-Vandeputte; Nicaise	Rich schools, poor schools. Hidden resource inequalities between primary schools	Y	Y	N	Y
58	Yalcin	Multi-level classification of literacy of educators using PIAAC data	Y	Y	N	Y
59	Kim et al.	Survey of secondary youth on relational aggression: impact of bullying social status, and attitudes	Y	N	Y	Y
60	Schultze-Krumbholz et al.	A Comparison of Classification Approaches for Cyberbullying and Traditional Bullying Using Data From Six European Countries	Y	N	Y	Y
61	Heiden-Rootes et al.	Peer Victimization and Mental Health Outcomes for Lesbian, Gay, Bisexual, and Heterosexual Youth: A Latent Class Analysis	Y	Y	Y	Y
62	Auer et al.	Multilevel Latent Class Analysis for Large-Scale Educational Assessment Data: Exploring the Relation Between the Curriculum and Students' Mathematical Strategies	Y	Y	N	Y
63	Pitzalis et al.	Cultural capital and educational strategies. Shaping boundaries between groups of students with homologous cultural behaviours	Y	Y	Y	Y
64	Marcoulides et al.	A latent transition analysis of academic intrinsic motivation from childhood through adolescence	Y	Y	Y	Y
65	Jeong et al.	Shifting Gears: Characteristics and Consequences of Latent Class Transitions in Doctoral Socialization	Y	Y	N	Y
66	Kaqinari et al.	A Latent Class Analysis of University Lecturers' Switch to Online Teaching during the First COVID-19 Lockdown: The Role of Educational Technology, Self-Efficacy, and Institutional Support	Y	N	Y	Y
67	Yamashita et al.	Adult Numeracy Skill Practice by STEM and Non-STEM Workers in the USA: An Exploration of Data using Latent Class Analysis	Y	Y	Y	Y

68	Myers et al.	Teacher Qualification Typologies and Their Relationship With the Math Achievement of Adolescents At Risk for Math Difficulties: A Latent Class Analysis Study	Y	Y	N	Y
69	Ing, Nylund-Gibson	Linking early science and mathematics attitudes to long-term science, technology, engineering, and mathematics career attainment: latent class analysis with proximal and distal outcomes	Y	Y	Y	Y
70	Ramos et al.	From College-to-Work: Latent Class Models Analysis of Mutual Adjustment in Internships after the Diploma	Y	Y	N	Y
71	Drossel, Eickelmann	Teachers' participation in professional development concerning the implementation of new technologies in class [...].	Y	Y	Y	Y
72	Weisz, Karim	Weisz communication styles inventory (WCSI: Version 1.0): development and validation	N	N	N	Y
73	Wu et al.	A new perspective on memorization practices among East Asian students based on PISA 2012	Y	Y	N	Y
74	Lewalter et al.	Investigating Visitor Profiles as a Valuable Addition to Museum Research	Y	N	N	Y
75	Zhang et al.	Teacher perceptions of effective professional development: insights for design	Y	Y	N	Y
76	Bofah, Hannula	Home resources as a measure of socio-economic status in Ghana	Y	Y	N	Y
77	Pohlenz et al.	How do students deal with forced digitalisation in teaching and learning? Implications for quality assurance	Y	N	Y	Y
78	McGrath et al.	Examination of college student health behaviors and self-reported executive functions	Y	N	Y	Y
79	Engledow et al.	Profiles of Elementary Teachers' Use of Mathematics Curriculum Materials and the Influence of Teacher Expertise	Y	Y	Y	Y
80	Gao et al.	Re-validating a Learning Progression of Buoyancy for Middle School Students: A Longitudinal Study	N	N	N	N

81	Garnett et al.	Coping Styles of Adolescents Experiencing Multiple Forms of Discrimination and Bullying: Evidence From a Sample of Ethnically Diverse Urban Youth	Y	Y	Y	Y
82	Penuel et al.	Teaching with student response systems in elementary and secondary education settings: A survey study	N	N	N	Y
83	Boyce, Bowers	Principal Turnover: Are There Different Types of Principals Who Move From or Leave Their Schools? A Latent Class Analysis of the 2007-2008 Schools and Staffing Survey and the 2008-2009 Principal Follow-Up Survey	Y	Y	Y	Y
84	Alexander, et al.	Racial/Ethnic Differences in Chronic Disease Predictors Among American High School Students	Y	Y	Y	Y
85	Kim, Fram	Profiles of choice: Parents' patterns of priority in child care decision-making	Y	Y	N	Y
86	Barringer, Jaquette	The Moving Missions of Community Colleges: An Examination of Degree-Granting Profiles Over Time	Y	Y	N	Y
87	McCulloch	Educational Aspirations Trajectories in England	Y	Y	N	Y
88	Yaman, Nerdel	Identification of student types based on their knowledge and their interests when learning with computer simulations	N	N	N	Y
89	Chen, Lin	A Cross-Cultural Study of Mathematical Achievement: from the Perspectives of One's Motivation and Problem-solving Style	Y	Y	Y	Y
90	Lamb et al.	Psychosocial factors impacting STEM career selection	Y	N	N	Y
91	Guo et al.	Cyberbullying Roles Among Adolescents: A Social-Ecological Theory Perspective	Y	Y	Y	Y
92	Henry et al.	Typologies of Stressful Life Events and Their Association With Sexual Risk Behaviors and Communication Among Justice-Involved Males and Their Female Sex Partners	Y	Y	N	Y
93	Sinclair et al.	Investigating Linguistically Diverse Adolescents' Literacy Trajectories Using Latent Transition Modeling	Y	Y	Y	Y

94	Askell-Williams et al.	Quality of implementation of a school mental health initiative and changes over time in students' social and emotional competencies	N	N	Y	Y
95	Fulmer et al.	Applying a Force and Motion Learning Progression over an Extended Time Span using the Force Concept Inventory	Y	N	N	Y
96	Park et al.	Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute	Y	Y	Y	Y
97	Yoon, Kim	Dynamic patterns of teachers' professional development participation and their relations with socio-demographic characteristics, teacher self-efficacy, and job satisfaction	Y	Y	Y	Y
98	Quirk et al.	Exploring patterns of Latino/a children's school readiness at kindergarten entry and their relations with Grade 2 achievement	Y	N	Y	Y
99	Hermansen	Danish Students? Use of ICT in Higher Education and its Perceived Meaningfulness	Y	Y	Y	Y
100	Lesterhuis et al.	Validity of Comparative Judgment Scores: How Assessors Evaluate Aspects of Text Quality When Comparing Argumentative Texts	Y	Y	N	Y
Total Counts (Y)			81	63	57	96

Appendix B

Article	Authors	Dichotomous	Indicator Number	Class Number
1	Burgos-Videla et al.	Y	8	4
2	Denson, Ing	Y	5	4
3	Liu et al.	Y	8	4
4	Choi, Kang	Y	14	6

5	Lingo, Chen	N	14	10
6	Kim et al.	N	12	3
7	Nasiopoulou et al.	Y	9	2
8	Custer, Akaeze	Y	18	5
9	Marraccini et al.	Y	6	4
10	Denson et al.	Y	5	3
11	Campbell et al.	Y	7	5
12	Akçakin, Kaya	Y	10	3
13	Berkowitz et al.	Y	7	4
14	Timmons, Pelletier	Y	6	4
15	Lounek, Ryska	N	9	3
16	Patte et al.	Y	5	4
17	Kang	Y	14	5
18	Khan, Krell	Y	7	2
19	Cai, Cheung	Y	12	3
20	Basaran, Yalman	Y	18	3
21	Urick	Y	18	4
22	Toker, Green	Y	10	3
23	Stein, Jimerson	N	4	3
24	Gao et al.	Y	14	5
25	Kang, Choi	Y	12	4

26	Greiff et al.	N	6	6
27	Brown	Y	12	2
28	Cho	Y	6	3
29	Agasisti et al.	Y	20	3
30	Karakoyun, Basaran	Y	21	4
31	Urick, Bowers	Y	11	3
32	Jones, Rosenberg	Y	9	4
33	Burns et al.	Y	11	4
34	Zhong et al.	Y	6	3
35	Weerts et al.	Y	8	4
36	Fematt et al.	Y	12	4
37	Murtafi'ah et al.	Y	9	2
38	Olivera-Aguilar et al.	Y	10	6
39	Postigo et al.	Y	5	3
40	Brandriet, et al.	N	3	5
41	King et al.	N	7	3
42	Fleary	Y	15	4
43	Bowers, Sprott	Y	20	4
44	Coombs et al.	Y	12	3
45	Choi, Kang	Y	11	5
46	Boutin-Martineet al.	Y	7	4

47	Duff, Bowers	Y	17	6
48	Dang, Nylund-Gibson	Y	8	2
49	Harshman, Yezierski	Y	7	7
50	Vaval et al.	Y	9	4
51	Harlow et al.	Y	9	2
52	Kim, D et al.	Y	8	4
53	Zhu, YT et al.	Y	8	3
54	Weissinger, et al.	Y	8	3
55	Helsabeck et al.	N	5	7
56	Gutiérrez et al.	Y	4	2
57	Poesen-Vandeputte; Nicaise	Y	8	3
58	Yalcin	N	3	3
59	Kim et al.	Y	14	3
60	Schultze-Krumbholz et al.	N	22	3
61	Heiden-Rootes et al.	Y	5	2
62	Auer et al.	Y	12	4
63	Pitzalis et al.	N	14	4
64	Marcoulides et al.	Y	4	3
65	Jeong et al.	Y	8	3
66	Kaqinari et al.	Y	8	4

67	Yamashita et al.	Y	12	4
68	Myers et al.	N	10	8
69	Ing, Nylund-Gibson	Y	10	4
70	Ramos et al.	N	3	2
71	Drossel, Eickelmann	Y	6	3
72	Weisz, Karim	N	15	4
73	Wu et al.	N	4	4
74	Lewalter et al.	Y	6	3
75	Zhang et al.	Y	12	4
76	Bofah, Hannula	Y	11	3
77	Pohlenz et al.	Y	12	3
78	McGrath et al.	Y	8	3
79	Engledowl et al.	Y	4	2
80	Gao et al.	Y	7	4
81	Garnett et al.	Y	5	3
82	Penuel et al.	N	5	4
83	Boyce, Bowers	Y	18	3
84	Alexander, et al.	Y	5	3
85	Kim, Fram	N	7	4
86	Barringer, Jaquette	N	4	5
87	McCulloch	N	4	6

88	Yaman, Nerdel	Y	5	4
89	Chen, Lin	Y	8	4
90	Lamb et al.	Y	10	3
91	Guo et al.	Y	8	4
92	Henry et al.	Y	17	4
93	Sinclair et al.	Y	5	3
94	Askell-Williams et al.	Y	37	2
95	Fulmer et al.	Y	17	4
96	Park et al.	Y	10	4
97	Yoon, Kim	Y	14	5
98	Quirk et al.	Y	16	5
99	Hermansen	Y	15	4
100	Lesterhuis et al.	Y	7	4
Mean (Range)		81	10 (3, 37)	4 (2, 10)

Appendix C

All associated code, models, and data to replicate the power simulation study and CLCA examples discussed in this thesis can be found in the following repository (https://github.com/garberadamc/SIM_CLCA). All analyses were conducted in R using the MplusAutomation package with models and simulation estimated by the Mplus program (R Core Team, 2017, Hallquist & Wiley, 2018, Muthén & Muthén, 2002). Appendix C.V includes the code to reproduce the response pattern tile plot (Figure 7). This Appendix is broken down into the following subsections:

1. Appendix C.I: Simulation
2. Appendix C.II: Split-Sample CLCA Method
3. Appendix C.III: Hypothesized-CLCA Method
4. Appendix C.IV: Replication Multi-Group Method
5. Appendix C.V: Plotting the Response Pattern Tile Plot