



Cite this: DOI: 10.1039/d5rp00432b

## Exploring heterogeneity in chemistry education research: comparing cluster analysis and latent profile analysis

Brandon J. Yik,<sup>a</sup> Yidi Zhang,<sup>b</sup> Karen Nylund-Gibson,<sup>c</sup> Marsha Ing,<sup>d</sup> Lillia Krawiec,<sup>c</sup> Joseph D. Houck<sup>e</sup> and Eric D. Nacsa<sup>e</sup>

Grouping approaches are commonly employed in chemistry education research to better understand variation. Traditionally used as a tool for data dimensionality reduction, these approaches are used as a tool to help researchers interpret complex data sets that can inform instructional strategies or target interventions. Among these techniques, cluster analysis, and in particular *k*-means clustering, has gained popularity for its simplicity and applicability to continuous variables. However, *k*-means cluster analysis is limited by its algorithmic nature, including assumptions of equal variance between clusters. Latent profile analysis, a model-based alternative within the mixture modeling framework, offers greater flexibility by allowing probabilistic group membership and the modeling of individual variances and covariances across latent profiles. This methods-focused study compares *k*-means clustering and latent profile analysis using data from undergraduate organic chemistry students enrolled in courses with either traditional or specifications grading. By examining students' affective traits, this study highlights the strengths and limitations of each grouping approach. Findings support the broader adoption of mixture modeling in chemistry education research to explore heterogeneity.

Received 21st November 2025,  
Accepted 6th April 2026

DOI: 10.1039/d5rp00432b

rsc.li/cerp

## Introduction

Grouping approaches based on a set of variables are a common strategy in chemistry education research, particularly for exploring the heterogeneity that exists within populations. For example, educational data often reflect students' diverse experiences, beliefs, and behaviors, which can be used to identify meaningful subgroups that can inform instruction and support. Traditionally, cluster analysis has been used to identify such groups, but model-based approaches like mixture modeling may afford greater flexibility and nuanced interpretation by accounting for variability in response patterns. This study uses an example from undergraduate organic chemistry to present a direct comparison between *k*-means cluster analysis and one type of mixture modeling, latent profile analysis LPA. By delineating the methodologies and statistical assumptions

of each approach, we aim to guide researchers in selecting the most appropriate grouping approach for their research goals, especially when addressing complex and heterogeneous educational data.

### Cluster analysis

Cluster analysis (or clustering) is a process of grouping observations, such as students, based on similarities in their data. Unlike classification methods, clustering does not rely on predefined group labels. Instead, clustering identifies natural groupings within the data. Clustering algorithms typically aim to optimize two criteria: (1) minimizing intracluster distance, which increases similarity between data within the same cluster and (2) maximizing intercluster distance, which increases dissimilarity between data in different clusters.

The goal of cluster analysis is to reduce a large number of data or observations into a smaller number of meaningful clusters that can be used to interpret patterns in the data (Everitt *et al.*, 2011). Clustering algorithms sort together similar or neighboring data points into clusters of data in *n*-dimensional space (Auf der Heyde, 1990). In chemistry education research, several types of clustering models have been used: connectivity models such as *hierarchical clustering* which group data based on similarity distances and are commonly used for categorical data (*e.g.*, Linenberger and Holme, 2014; Raker and

<sup>a</sup> Department of Chemistry, University of Georgia, Athens, GA 30602, USA.  
E-mail: byik@uga.edu

<sup>b</sup> Department of Communication, University of California, Santa Barbara, Santa Barbara, CA 93106, USA

<sup>c</sup> Department of Education, University of California, Santa Barbara, Santa Barbara, CA 93106, USA

<sup>d</sup> School of Education, University of California, Riverside, Riverside, CA 92521, USA

<sup>e</sup> Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA



Holme, 2014; Galloway and Bretz, 2015a, 2015b; Raker *et al.*, 2015a, 2015b; Nielsen and Yeziarski, 2016; Velasco *et al.*, 2016; Gibbons *et al.*, 2018, 2022; Lewis, 2018; Ferreira and Lawrie, 2019; Jeffery and Bauer, 2020; Popova *et al.*, 2020, 2021; Schultz *et al.*, 2021; Gulacar *et al.*, 2022); *centroid models* such as *k*-means clustering where data are grouped based on the mean of observations, where each cluster is represented by a single mean vector (*e.g.*, Jurišević *et al.*, 2012; Lastusaari and Murtonen, 2013; Guerris *et al.*, 2020; An *et al.*, 2022; Guo *et al.*, 2022; Wang and Lewis, 2022; Lee and Guo, 2024; Partanen *et al.*, 2024; Sizemore *et al.*, 2024; Braun *et al.*, 2025; Jaison *et al.*, 2025); and *distributions models* such as DBSCAN (density-based spatial clustering of applications of noise) where data are grouped based on how closely packed they are in space (*e.g.*, Martin *et al.*, 2024). Additionally, hybrid algorithms or sequential clustering approaches (*e.g.*, hierarchical clustering followed by *k*-means clustering) have also been used to refine cluster solutions (Brandriet and Bretz, 2014; Connor *et al.*, 2021; Wilkes *et al.*, 2024). Within chemistry education, hierarchical clustering methods are frequently used for categorical data, and *k*-means clustering is among the most common for numerical data.

***k*-Means cluster analysis.** In *k*-means clustering, each observation is partitioned into a pre-specified number of *k* nonoverlapping subgroups called clusters. The algorithm begins with an initial set of cluster assignments by randomly assigning each observation to one of the *k* clusters. For each cluster, the mean vector of all observations is computed, which serves as the cluster center or centroid. The algorithm then calculates the distance between each observation and each centroid reassigning observations to the cluster with the smallest distance to its centroid. Most commonly, Euclidean distance is used, although other distance metrics (*e.g.*, Manhattan or cosine distance) can be used depending on research context and data properties. This process continues iteratively to maximize the within-cluster sum of squares (WSS), or the sum of squared distances between each observation and its assigned cluster centroid.

The goal of *k*-means cluster analysis is to maximize the similarity between observations within the same cluster while minimizing dissimilarity between observations in different clusters. For example, if a researcher tests a range of cluster solutions from  $k = 2$  to  $k = 10$ , the algorithm repeats this process for each value of *k*, seeking to minimize the WSS for each solution. The final cluster solution is typically selected based on criteria such as the elbow method which identifies the point at which adding more clusters yields diminishing returns in reducing WSS.

### Applications of cluster analysis in chemistry education research

Cluster analysis has been widely applied to support student success in chemistry (see Auf der Heyde (1990) for an introduction of cluster analysis tailored to chemistry education researchers). Researchers have found value in identifying groups of students with similar characteristics, such as affect or experiences (Lewis *et al.*, 2009; Brandriet and Bretz, 2014; Chan and Bauer, 2014; Nielsen and Yeziarski, 2016; Jeffery and

Bauer, 2020; Jaison *et al.*, 2025). For example, Galloway and Bretz (2015a) used cluster analysis on data collected from undergraduates in a first-year chemistry laboratory course using the Meaningful Learning in the Laboratory Instrument (MLLI) to identify students who were similar in terms of their expectations and experiences. The four clusters identified suggested that students' expectations about the course were related to their experiences in the course. Students who were in the cluster identified by the researchers as "low" leaned toward having more negative affective expectations for their experiences, and showed decreases across cognitive, affective, and cognitive/affective scales after experiencing the course. In contrast, students who were in the cluster identified as "high" had higher reported affective and cognitive experiences, but saw a decrease in their cognitive scores, increase for the affective scale, and no change for the cognitive/affective scale. This research highlighted how identifying clusters of students can help instructors better understand and address varying expectations in laboratory courses.

Beyond student-focused research, cluster analysis has been used to group instructors (*e.g.*, van Driel *et al.*, 2005; Drechsler and Van Driel, 2009; Linenberger and Holme, 2014; Gibbons *et al.*, 2018) and course materials (*e.g.*, Raker *et al.*, 2015b; Ferreira and Lawrie, 2019). Researchers have applied cluster analysis on data from multiple data sources (*e.g.*, surveys and interview data) and across different samples to examine changes over time, such as shifts in faculty beliefs (Popova *et al.*, 2020, 2021).

Chemistry education researchers have also related clusters to other variables to explore theoretical relationships and assumptions about learning (*e.g.*, Raker and Holme, 2014). For example, An *et al.* (2022) investigated whether high failure rates in introductory chemistry courses were associated with students' study approaches. They found that students within clusters characterized by "ideal" studying approaches had higher exam scores and final course grades, suggesting that clustering can reveal meaningful patterns in student behavior and outcomes.

While cluster analysis has proven valuable in chemistry education research, several limitations should be considered, particularly with the use of *k*-means clustering, one of the most commonly applied algorithms. A recurring concern is that clustering solutions are sometimes presented with limited theoretical justification for why a particular clustering algorithm is appropriate for the research question and data structure. Although *k*-means clustering does not assume equal cluster sizes, it assigns observations to the nearest centroid under the chosen distance metric to minimize within-cluster dispersion, which can result in clusters of very different sizes. Additionally, the algorithm requires the number of clusters to be specified in advance, often without a model-based rationale, which can lead to arbitrary or unstable solutions. This clustering algorithm is also sensitive to initial starting values and converges on local optima, potentially leading to inconsistent cluster assignments. Also, reliance on distance-based assignment can be influenced by scaling and outliers. Finally, because



$k$ -means provides deterministic (“hard”) assignments rather than probabilistic membership estimates, it offers limited information about classification uncertainty relative to model-based approaches. These limitations echo long-standing concerns by quantitative methodologists (Blashfield and Aldenderfer, 1988). Mixture modeling (described in next section) addresses several of these known limitations with cluster analysis by offering a probabilistic, model-based framework for estimating latent subgroups and evaluating competing solutions.

### Latent profile analysis (LPA)

Mixture models are a family of statistical methods within the broader latent variable modeling framework (Collins and Lanza, 2010). The primary goal of mixture models is to identify unobserved (*i.e.*, latent) subgroups in a population based on patterns of responses to observed variables. One type of mixture model is latent profile analysis (LPA), which is appropriate when the observed variables (*i.e.*, *indicators*) are continuous.

The commonly used latent variable models (*e.g.*, factor analysis or item response theory) which estimate *continuous* latent variables and aim to group variables and thus can be referred to as variable-centered approaches. LPA, however, estimates a *categorical* latent variable that groups individuals into mutually exclusive profiles based on their response patterns. Because LPA groups people rather than variables, it is sometimes referred to as a person-centered approach. This makes LPA particularly useful for identifying qualitatively distinct subpopulations within a heterogeneous sample.

**Model estimation.** LPA identifies latent profiles within a population based on individuals’ pattern of responses to continuous indicators. LPA models are typically estimated using robust maximum likelihood estimation which uses an iterative process called the expectation–maximization (EM) algorithm. During the *initialization step*, the observations are divided into  $k$  classes that are equal in size. Then, in the *expectation step*, the algorithm computes the probability that an observation belongs to each of the  $k$  latent profiles by computing the posterior probabilities of profile assignment using the current model parameter estimates (*i.e.*, relative profile size and conditional item probabilities). Next, in the *maximization step*, the algorithm updates the model parameters to improve model fit by maximizing the expected log-likelihood function *via* updating the model parameter estimates based on the expected parameter values obtained in the expectation step. The expectation and maximization steps are repeated iteratively until the log-likelihood function converges on a solution that provides  $k$  number of latent profiles with the size of each profile and the conditional probabilities for each item in each profile.

**Profile enumeration.** Determining the optimal number of latent profiles is known as *profile enumeration*. This process involves fitting a series of models with an increasing number of profiles and comparing model fit information to determine the appropriate model among those considered. In this process, a one-profile model serves as the comparative baseline for models with more than one profile. Iteratively increasing the number of profiles by one allows for comparison with

the preceding model. The model fit can be compared through several fit indices which include information criteria (*e.g.*, AIC, BIC, adjusted BIC) which balance model fit and complexity, likelihood ratio tests (*e.g.*, Lo-Mendell-Rubin LRT, Bootstrap LRT) which test whether adding an additional profile improves model fit, and the Bayes Factor (BF) and correct model probability (cmP) which provide probabilistic comparisons across models. More details on fit indices and model selection are provided in the Methods section and in works by Nylund *et al.* (2007), Masyn (2013), and Nylund-Gibson and Choi (2018).

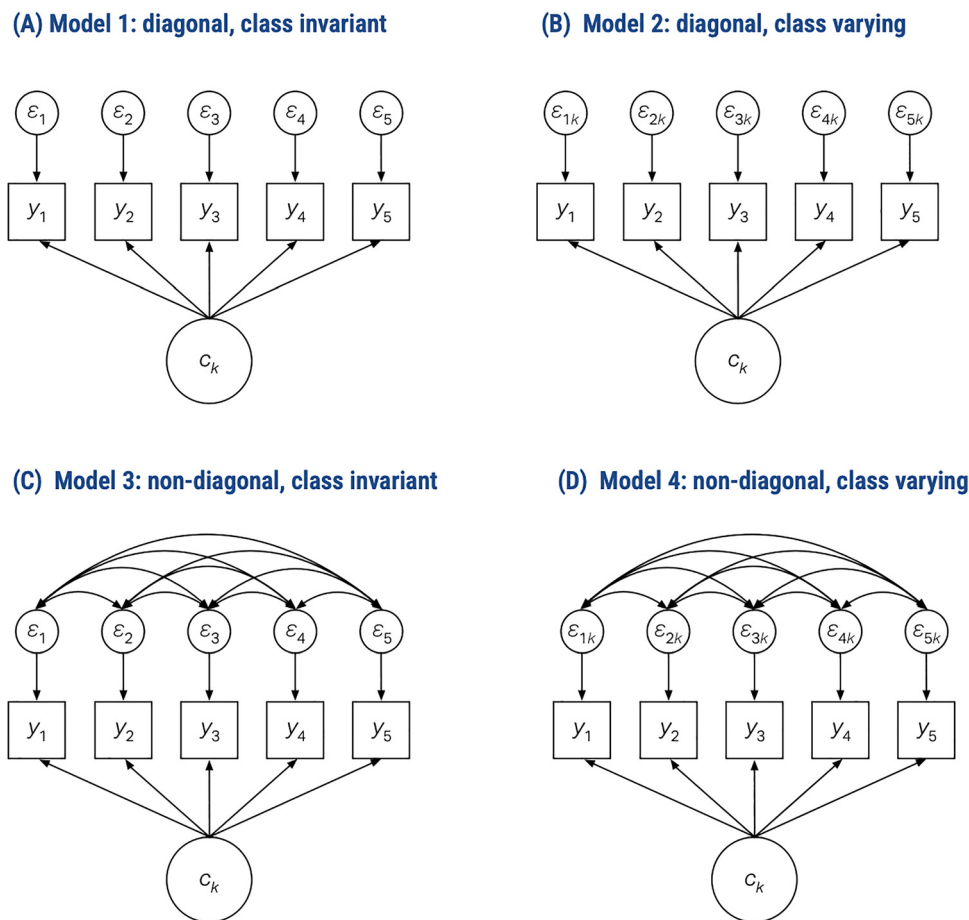
Although fit indices provide valuable guidance, selecting the number of profiles is ultimately an inferential judgment that balances statistical evidence, parsimony, theoretical coherence, and interpretability. In practice, difference indices may point to different solutions, and tools that aggregate multiple indices (*e.g.*, *NbClust* for clustering; Charrad *et al.*, 2014) can support, but not replace, researcher judgment.

**Model specifications.** Due to the continuous nature of the indicators, LPA requires the consideration of a range of model specifications relating to the item (*e.g.*, item means, item variances, and item covariances). This added complexity introduces additional models that must be considered during the profile enumeration process (Peugh and Fan, 2013). Specifically, four different models must be considered in LPA, each differing in how item variances and covariances are treated. Fig. 1 depicts the path diagram for these four models with five indicators, as an example:

- Model 1: *diagonal, class invariant*: item variances are constrained to be equal across profiles and item covariances are not estimated (Fig. 1A).
- Model 2: *diagonal, class varying*: item variances are freely estimated across profiles, and item covariances are not estimated (Fig. 1B).
- Model 3: *non-diagonal, class invariant*: item variances are constrained to be equal across profiles, and item covariances are estimated (Fig. 1C).
- Model 4: *non-diagonal, class varying*: item variances are freely estimated across profiles, and item covariances are estimated (Fig. 1D).

In these models, *diagonal* refers to the absence of item covariances (*i.e.*, zero correlations among indicators), *non-diagonal* allows for the estimation of item covariances, *class-invariant* means items variances are held equal across profiles, and *class-varying* allows item variances to differ across profiles. Although the literature often uses the term “class” in describing these models, this manuscript uses the term “profile” to refer specifically to the latent groups identified *via* LPA. Specifically, Model 1 assumes that the indicators are independent within profiles and have equal variances across profiles. Model 2 allows indicator variances to differ across profiles while maintaining the assumption of no within-profile covariances. Model 3 constrains variances to be equal across profiles but allows indicators to covary within profiles. And, Model 4 is the most flexible specification, allowing both indicator variances and covariances to vary across profiles.





**Fig. 1** Latent profile enumeration models. (A) Model 1: diagonal, class invariant; (B) Model 2: diagonal, class varying; (C) Model 3: non-diagonal, class invariant; (D) Model 4: non-diagonal, class varying.  $c_k$  represents the categorical latent grouping variable  $c$  with  $k$  groups.  $y_1$  through  $y_5$  represent the item means of five continuous indicators with corresponding  $\varepsilon_1$  through  $\varepsilon_5$  item variances that are equal across groups and  $\varepsilon_{1k}$  through  $\varepsilon_{5k}$  representing item variances that vary across groups.

### Applications of LPA in chemistry education research

To date, only a handful of studies have applied LPA in chemistry education. Although LPA appears less frequently than traditional clustering approaches, existing work suggests that it can yield interpretable profiles related to student understanding, instructional practices, and affective outcomes. For example, Villalta-Cerdas and Sandi-Urena (2014) used LPA to identify profiles based on students' explanations of entropy and the second law of thermodynamics. Stains *et al.* (2018) applied LPA to observed classroom behaviors to determine instructional profiles among science, technology, engineering, and technology (STEM) faculty. Hensen and Barbera (2019) used LPA to group students based on affective response to general chemistry laboratory experiences (*e.g.*, emotional satisfaction, intellectual accessibility, and equipment usability). Additionally, Liu and colleagues (2026) used LPA to identify profiles of Chinese high school students' chemistry achievement goals. More recently, Pulukuri *et al.* (2024) expanded the use of LPA by identifying latent profiles of self-efficacy among general and organic chemistry students and relating profile membership to distal variables (*e.g.*, exam performance), and covariates (*e.g.*, student identities and course level).

Despite this promise, chemistry education applications do not always capitalize on LPA's flexibility in specifying within-profile (co)variance structures, which can meaningfully affect solutions and their interpretation. In many LPA workflows, particularly those implemented in software where the typical starting point is a class-invariant, diagonal model, researchers often begin with (and sometimes only report) the highly constrained specification that fixes indicator covariances to zero and constrains variances to equality across profiles (often described as the "default" LPA model in Mplus-style parameterizations; Muthén and Muthén, 2017). However, this pattern is not universal across software. For example, R's *mclust* framework fits Gaussian mixture models across a range of covariance structures and selects an optimal model using one metric (typically the Bayesian Information Criterion, BIC; Scrucca *et al.*, 2023), and packages such as *tidyLPA* can support comparing multiple profile specifications (Rosenberg *et al.*, 2018).

Thus, many published chemistry education studies using LPA either (a) rely on a constrained diagonal specification or (b) do not clearly report evaluating alternative variance/covariance structures, even though methodological guidance recommends



systematically considering such alternatives when warranted by theory and data (Masyn, 2013; Nylund-Gibson and Masyn, 2016; Nylund-Gibson and Choi, 2018). Considering these alternative models is important because misspecifying within-profile and/or covariances can distort profile enumeration and interpretation, potentially leading to the over- or underestimation of the right number of latent profiles.

Although LPA has been used less frequently in chemistry education than methods such as *k*-means clustering, the more important concern is not its limited use, but how it is often implemented and reported. In particular, studies frequently begin with simple model specifications and do not transparently report whether alternative model structures were considered. This lack of transparency can often make it difficult to evaluate the validity and interpretability of the resulting groupings.

To address this issue, the present study explicitly reports key LPA modeling decisions, including assumptions about within-profile variances and covariances, and compares the resulting subgroup solutions with those obtained from *k*-means clustering. By examining where the two approaches converge or differ, this study aims to support more rigorous and interpretable subgrouping research in chemistry education.

### Research background and goals

In this study, we describe the process of applying both *k*-means cluster analysis and LPA to a data set relevant to chemistry education to compare their processes and outcomes. Each analysis is conducted using the current best practices for its respective technique. Our goal is to compare and contrast the methodological assumptions, strengths, and limitations of *k*-means clustering and LPA in the context of this data set, providing practical guidance for researchers selecting grouping approaches in chemistry education research.

The data used is relevant to chemistry education and measures alternative grading. Extensive empirical evidence has shown that traditional grading schemes (*e.g.*, A–F grades or percentage-based scales) often fail to accurately reflect student learning and can reinforce systemic inequities in education (Matz *et al.*, 2017; Feldman, 2019a, 2019b; Link and Guskey, 2019; Clark and Talbert, 2023). These disparities are rooted in unequal access to resources, opportunities, and support, and are embedded within the structure and policies of our current education system (Patton, 2016; Renn and Reason, 2023). In response, grading reforms have emerged over the last few decades, collectively referred to as the “alternative grading movement” (Clark and Talbert, 2023; Hackerson *et al.*, 2024). Alternative grading systems such as ungrading (*e.g.*, Blum, 2020), standards-based grading (*e.g.*, Boesdorfer *et al.*, 2018), competency-based grading (*e.g.*, Diegelman-Parente, 2011), contract grading (*e.g.*, Offerdahl *et al.*, 2016), labor-based grading (*e.g.*, Inoue, 2022), and specifications grading (*e.g.*, Nilson, 2015; Howitz *et al.*, 2021; Nilson and Packowski, 2026) have gained attention across STEM disciplines. In chemistry, specifications grading, or “specs grading” has emerged as the most prominent alternative grading model (Hackerson *et al.*, 2024;

Wang *et al.*, 2025b), with several works describing its implementation and foundations (Nilson, 2015; Yik *et al.*, 2024, 2025; Wang *et al.*, 2025b; Nilson and Packowski, 2026).

Specifications grading is hypothesized to benefit both instructors and students (Nilson, 2015; Nilson and Packowski, 2026). For instructors, specifications grading is proposed to simplify grading, save time, uphold high academic standards, and improve interrater reliability. For students, it is expected to increase motivation, clarify expectations, reduce stress, and discourage academic dishonesty. While some studies have begun to empirically evaluate these claims (*e.g.*, Howitz *et al.*, 2021; Hunter *et al.*, 2022; Noell *et al.*, 2023; Closser *et al.*, 2024; Yik *et al.*, 2024, 2025), there remains a need for deeper investigations into the affective outcomes associated with specifications grading.

Guided by these hypothesized benefits of specifications grading as a framework, this study explores its potential impact on students’ affective experiences. While Nilson (2015) highlights increased student motivation as a central target, other affective constructs, such as attitudes, self-concept, effort beliefs, and self-efficacy, may also be impacted. Prior work in chemistry has empirically examined the effect of specifications grading on students’ motivation (Yik *et al.*, 2024), but to our knowledge, no study to date has leveraged a person-centered, multivariate approach to identify combinations of affective beliefs and examine how these groupings relate to student outcomes.

To justify the selection of affective indicators used for grouping, we ground each construct in well-established motivation and learning frameworks with strong empirical traditions in education. First, because specifications grading is intended to shift students’ reasons for engaging in coursework (*e.g.*, from point accumulation toward learning and mastery), we included motivation using self-determination theory, which distinguishes more autonomous and controlled motivation and explicitly links instructional contexts to the quality of students’ motivation (Deci and Ryan, 1985, 2000; Ryan and Deci, 2000). Second, because specifications grading also targets students’ expectations for success and the perceived value of course assessments (*via* transparent criteria and mastery-oriented standards), we included self-efficacy and attitudes/value-related evaluations as core constructs emphasized in expectancy–value perspectives and social-cognitive theory, both which predict persistence, engagement, and achievement (Bandura, 1977; Eccles and Wigfield, 2002). In addition, we included perceived competence/self-concept alongside self-efficacy because prior work demonstrates that these constructs are related but meaningfully distinct in definition and function (*e.g.*, broader self-evaluations *versus* task- and context-specific capacity beliefs), making them potentially separable components of students’ affective experience under a new grading system (Bong and Skaalvik, 2003). Third, we included attitudes toward chemistry because this construct has a strong measurement base in chemistry education and captures affective and value-related responses that are plausibly influenced by grading systems that change feedback, evaluation, and assessment (Brandriet *et al.*, 2011).



Finally, we included effort beliefs because specifications grading emphasizes revision, mastery, and meeting standards, which may alter how students interpret effort (*e.g.*, effort as productive *versus* evidence of low ability). This rationale is supported by research on implicit theories and related attributional processes showing that beliefs about effort and ability are linked to persistence and responses to difficulties (Mueller and Dweck, 1998; Blackwell *et al.*, 2007). Taken together, these theory-anchored constructs provide a principled and chemistry-relevant basis for modeling heterogeneous affective groupings and examining how different patterns of beliefs relate to student outcomes.

In this study, we apply both *k*-means clustering and LPA to explore the following research goals using data from undergraduate organic chemistry students:

- (1) To what extent do *k*-means clustering and LPA produce different affective groupings of students at the beginning of an organic chemistry course?
- (2) How do the affective groupings identified by each method relate to students' standardized exam performance?

## Methods

### Ethical considerations

This study was conducted under STUDY00021800 as reviewed and determined to be “research exempt” by The Pennsylvania State University Institutional Review Board. Informed consent was collected on paper during the first day of class by personnel who were not associated with the course nor study, and who deidentified data prior to analysis for the researchers. Authors JDH and EDN played a dual role as instructors of the course, but their role was to make students aware of this study and incentivize students (more details below). These authors were never made aware which students participated in this study.

### Institutional context

This study was conducted at a large public research university in the mid-Atlantic region of the United States. As of Fall 2024, this institution enrolled over 42 000 undergraduate students. Men and women comprise 52% and 47% of the undergraduate population respectively, and less than 1% are students who hold other gender identities. This university is predominately white, enrolling 20% of students who are U.S. residents of an ethnicity that is federally considered of underrepresented minority status; these students identify as American Indian/Alaska Native, Black/African American, Hispanic/Latino, Native Hawaiian/Pacific Islander, or as two or more combinations of these racial/ethnic categories. Approximately 9% of undergraduates are international students and 17% are first-generation students.

### Participants and data collection

Study participants were undergraduate students enrolled in one of two sections of a first-semester organic chemistry course during Fall 2024. One course section implemented a traditional grading scheme, while the other used specifications grading.

Students were invited to participate in the study at the beginning of the semester *via* an email invitation that contained a link to a Qualtrics survey. After listwise deletion for incomplete responses, the consent rate was 61.4% for the traditionally-graded course ( $n = 129$ ;  $N = 210$ ), and 66.4% for the specifications-graded course ( $n = 166$ ;  $N = 250$ ).

To encourage participation, students were incentivized regardless of whether they consented to participate in the study. In the specifications-graded course, students who completed the survey received 15 XP (experience points), which are additional credits earned and allow for modification of learning target quizzes (see syllabus in the SI). In the traditionally-graded course, students received 1 bonus point on the Foundations module. Detailed descriptions of the grading implementations for both courses are available through the course syllabi in the SI.

### Study measures

Five latent traits were of interest: attitudes, motivation, effort beliefs, self-efficacy, and self-concept. To appropriately measure these traits, we made minor wording changes and adjustments to the measurement scales as needed (Ferrell and Barbera, 2015) and are described below for each measure, referred to as indicator variables henceforth. Participants were given these modified scales *via* an online survey. Only complete response sets were included in the following analyses: data were listwise-deleted if participants did not answer all items for all five measures. Additionally, participants were asked to report demographic and contextual information, including their year in college, credit load, employment or internship status, first-generation status, gender identity, and racial/ethnic identity. The full survey can be found in the SI. Since LPA assumes that observed covariation among indicators is explained by the latent profile variable, traditional measures of validity, such as those used for factor analysis, are neither required nor diagnostic for evaluating LPA models.

**Chemistry motivation.** Motivation was measured using the Academic Motivation Scale–Chemistry (AMS-C; Liu *et al.*, 2017), a 28-item instrument designed to assess extrinsic motivation and intrinsic motivation in a way that is aligning with self-determination theory (Deci and Ryan, 2000; Ryan and Deci, 2000). The AMS-C is an adaptation of the broader Academic Motivation Scale (Vallerand *et al.*, 1992) and has been used in multiple chemistry contexts: general chemistry (Liu *et al.*, 2017), organic chemistry (Liu *et al.*, 2018), and inorganic chemistry (Pratt *et al.*, 2023). In line with previous chemistry-specific studies (Raker *et al.*, 2019; Pratt and Raker, 2020; Pratt *et al.*, 2023), the question stem was modified to make the AMS-C more course-specific: “*To what extent each of the following statements corresponds to one of the reasons why you are enrolled in this organic chemistry course.*” The original 5-point Likert scale ranging “*not at all*” to “*exactly*” was retained.

**Chemistry attitudes.** Attitudes were measured using the Attitude toward the Subject of Chemistry Inventory version 2 (ASCI V2; Xu and Lewis, 2011). The original 20-item instrument (ASCI V1) measured attitude with five subscales (*i.e.*, interest



and utility, anxiety, intellectual accessibility, fear, and emotional satisfaction) (Bauer, 2008), and was later revised as an 8-item short form (ASCI V2) to measure attitude with two subscales, intellectual accessibility and emotional satisfaction, which are indicators of cognition and affect, respectively (Xu and Lewis, 2011). The original 7-point semantic differential measurement scale was retained.

**Chemistry self-concept.** Self-concept was measured using the 10-item self-concept scale of the Chemistry Self-Concept Inventory (CSCI; Bauer, 2008). The CSCI asked participants, “How accurately does each statement describe you?” Like the MSLQ, the measurement scale for the CSCI also used a 7-point Likert scale with seven numerical choices with “1” representing “very inaccurate” and “7” representing “very accurate.” To remain consistent in the number of response options for measures with Likert scales, we also condensed the number of response options to a 5-point Likert scale.

**Effort beliefs.** Effort beliefs were measured using a 9-item scale originally developed by Sorich and Dweck (1997) and later used in a dissertation study by Blackwell (2002). The effort beliefs scale was then adapted for use in high school mathematics (Jones *et al.*, 2012), and later for use in postsecondary general chemistry laboratory (Ferrell and Barbera, 2015), which we use in this study. We retain Ferrell and Barbera’s (2015) adjusted measurement scale range of a 5-point Likert scale ranging from “strongly disagree” to “strongly agree.”

**Self-efficacy.** Self-efficacy was measured using the 8-item self-efficacy scale of the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich and De Groot, 1990). The MSLQ asked participants to use the measurement scale to answer the questions. The measurement scale used is a 7-point Likert scale with seven numerical choices; at one end, “1” represents “not at all true of me” and “7” represents “very true of me.” For clarity, we modified the 7-point Likert scale to a 5-point Likert scale with the same numerical descriptors to align with the other measurement scales used in this study. Additionally, previous work has shown that condensing the number of response options has little effect on the distribution about the mean, skewness, or kurtosis (Dawes, 2008).

**ACS exam score.** A custom first-semester American Chemical Society (ACS) organic chemistry exam was administered as the final exam for both course sections. The exam consisted of 50 multiple-choice questions. Topics on the exam included: structure and bonding, nomenclature, charge stability and resonance, molecular orbital theory, conformation of alkanes and cycloalkanes, acid–base chemistry, organic reactions and mechanisms (*e.g.*, substitution, elimination, electrophilic addition, and radical substitution), mass spectrometry, infrared spectroscopy, and nuclear magnetic resonance spectroscopy.

### ***k*-Means cluster analysis**

*k*-Means cluster analysis is a non-hierarchical, partition-based clustering algorithm that assigns observations to a pre-specified number of clusters in a way that minimizes within-cluster variance (MacQueen, 1967). *k*-Means clustering was conducted in R using the base implementation (R Core Team, 2025).

We classified students into latent profiles based on five *z*-score standardized indicator variables: attitudes toward chemistry, motivation to learn chemistry, effort beliefs, self-efficacy, and chemistry self-concept.

To determine the optimal number of clusters, we tested solutions ranging from one to ten clusters. For each solution, the clustering algorithm was run with a maximum of 25 iterations, and 10 random starts to reduce the risk of convergence on a local minimum. A fixed random seed (123) was used to ensure reproducibility.

Model fit was evaluated using multiple model fit indices, including the total within-cluster sum of squares (WSS), the average silhouette width which is a measure of how well-separated the clusters are, and the elbow method which is a visual inspection of a scree plot of WSS values across cluster solutions to identify the point at which adding more clusters yields diminishing returns. The optimal number of clusters was selected by identifying the elbow point in the WSS plot and considering solutions with higher silhouette coefficients which indicate better-defined and more distinct clusters.

### **Latent profile analysis**

Latent profile analysis (LPA) was used to identify different profiles of students’ affect in an organic chemistry course. We classified students into latent profiles based on the same five *z*-score standardized indicator variables: attitudes toward chemistry, motivation to learn chemistry, effort beliefs, self-efficacy, and chemistry self-concept.

**Latent profile enumeration.** Profile enumeration was conducted using recommended methods for class-invariant and class-varying diagonal models (Masyn, 2013; Nylund-Gibson and Masyn, 2016; Nylund-Gibson and Choi, 2018). Specifically, we evaluated LPA models under the four variance–covariance specifications previously described. These model specifications allow a comprehensive examination of the latent profile structure, ranging from highly restricted and parsimonious to unrestricted and more flexible.

LPA analyses were performed using the *tidyLPA* package (Rosenberg *et al.*, 2018) in R (R Core Team, 2025) to illustrate the utility of a free software package. We tested a series of iterative latent profile models based on our selected indicator variables and examined models with up to six groups (*e.g.*, six profiles) for each of the four specifications based on variance–covariance differences. Model fit information for each model was tabulated and used to determine the best-fitting solution. To reduce the risk of convergence on local maxima, models were estimated using multiple sets of random starting values (STARTS = 500, 100) and verified that the best log-likelihood solution was replicated across runs.

Although there is no consensus on the minimum sample size requirements for LPA, prior research suggests that larger sample sizes ( $n > 500$ ) are more likely to accurately identify subgroups (Hickendorff *et al.*, 2018). However, empirical applications with smaller sample sizes (*e.g.*,  $n = 150$ ) can be carried out appropriately if both the number of indicator variables and the number of latent profiles are relatively low



(*e.g.*, Nylund-Gibson and Choi, 2018; Gaias *et al.*, 2019; Knox *et al.*, 2025).

**Model fit evaluation.** Several comparative model fit indices were used to determine the best-fitting unconditional model. The first set of fit indices used are the information criteria (IC): the Bayesian Information Criterion (BIC), Consistent Akaike's Information Criterion (CAIC), adjusted Bayesian Information Criterion (aBIC), Approximate Weight of Evidence Criterion (AWE; Nylund *et al.*, 2007). The information criteria (*i.e.*, CAIC, BIC, aBIC, and AWE) evaluate model fit by accounting for sample size and the number of estimated parameters. Models with the lowest criterion values are generally considered to offer the best fit (Burnham *et al.*, 2011).

The second set of fit indices used are the likelihood ratio tests: Vuong-Lo-Mendell-Rubin (VLMR), and Bootstrapped Likelihood Ratio Test (BLRT; Nylund *et al.*, 2007). Unlike in structural equation modeling, where chi-squared different tests are used, these likelihood-based tests (*i.e.*, VLMR and BLRT) evaluate whether adding an additional profile significantly improves model fit, where a non-significant *p*-value (*i.e.*,  $p > 0.05$ ) suggests that the additional profile does not enhance model fit (Nylund *et al.*, 2007; Nylund-Gibson and Choi, 2018). For example, these tests compare the fit between two neighboring class/profile models (*e.g.*, a two-class/profile model *vs.* a three-class/profile model).

Two additional fit indices are used from the Bayesian framework: the Bayes Factor (BF) and the correct model probability (cmP). The BF provides a pairwise comparison between two neighboring models and the cmP provides an estimate of each estimated model being correct out of all the estimated models, assuming that the "true" model was estimated (Nylund *et al.*, 2007). Both the BF and cmP include the BIC in their calculation and are often highly correlated with BIC (Nylund *et al.*, 2007).

In addition to these statistical criteria, we considered how the emergent classes were supported by theoretical consideration and examined the proportion of individuals within each profile to guide model selection (Nylund-Gibson and Choi, 2018).

**Latent class regression.** After identifying the best-fitting profile solution, we applied the automated three-step approach implemented in the *tidyLPA* package (Rosenberg *et al.*, 2018) to examine the relationships between latent profile membership, a predictor (*i.e.*, grading scheme), and a distal outcome

(*i.e.*, ACS exam score) in a method called latent profile regression. Using this approach after deciding on the number of profiles provides more insight into the students who comprise each of the profile groups.

This manuscript demonstrates using the free *tidyLPA* package to implement the three-step approach for analyzing the relationships between latent profiles and auxiliary variables. This approach begins by estimating the unconditional LPA model independently of auxiliary variables (*e.g.*, predictors and distals). In the second step, individuals are assigned to profiles based on their most likely profile membership. In the third step, profile membership is treated as an observed variable to examine associations with auxiliary variables.

It is important to note that, unlike the manual maximum likelihood (ML) three-step method (Asparouhov and Muthén, 2014), the *tidyLPA* package does not adjust for classification error, and instead assumes perfect classification. Assuming perfect classification in LPA can be problematic because it ignores the uncertainty in class membership, leading to biased estimates in auxiliary analyses, inflated distinctions among profiles, and underestimated standard errors, which can result in misleading conclusions (Vermunt, 2017). Following profile assignment, omnibus tests were conducted to assess overall associations between profiles and each auxiliary variable. When omnibus tests were statistically significant (*i.e.*,  $p < 0.05$ ), pairwise comparisons were performed to determine specific differences between profiles.

### Comparing *k*-means cluster analysis and latent profile analysis

After determining solutions for each method, we systematically compared the resulting classification across three dimensions: (1) the consistency of group membership across the two methods, (2) the similarity of group characteristics including profile means and variances, and (3) impact on distal outcomes, specifically on standardized exam performance.

## Results and discussion

### Descriptive statistics

To provide an overview of the study variables, descriptive statistics and bivariate correlations are reported in Table 1. The five continuous indicators of students' affect include

Table 1 Indicator variable descriptive statistics and correlations

	<i>M</i> (SD)	1	2	3	4	5	6	7
1. Chemistry motivation	3.07 (0.53)	1.00						
2. Chemistry attitudes	3.68 (0.57)	−0.27***	1.00					
3. Chemistry self-concept	2.83 (0.35)	0.16**	0.10	1.00				
4. Effort beliefs	2.89 (0.42)	0.02	0.14*	0.34***	1.00			
5. Self-efficacy	3.57 (0.84)	0.22***	−0.15*	−0.12	−0.37***	1.00		
6. Grading scheme	0.56 (0.50)	−0.09	0.06	0.12*	0.07	0.11	1.00	
7. ACS exam score	66.79 (14.04)	0.03	−0.16*	−0.02	−0.21***	0.37**	0.12	1.00

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Chemistry motivation, chemistry self-concept, effort beliefs, and self-efficacy were measured on 5-point scales. Chemistry attitudes were measured on a 7-point scale. Grading scheme = 0 represents traditional grading and grading scheme = 1 represents specifications grading. ACS exam score was out of 100 points.



chemistry motivation, chemistry attitudes, chemistry self-concept, effort beliefs, and self-efficacy. In addition to these indicators, two auxiliary variables were included to support the interpretation of both the *k*-means cluster analysis and LPA results: grading scheme and ACS exam score. Grading scheme (*i.e.*, traditional *vs.* specifications grading) was treated as a covariate, serving as a predictor of subgroup membership. ACS exam score was treated as a distal outcome, used to how subgroup membership relates to standardized exam performance.

### *k*-Means cluster analysis

To determine the optimal number of clusters, we examined both the total within-cluster sum of squares (WSS) and the average silhouette width across solutions ranging from 1 to 10 clusters. The WSS sharply decreased from one to two clusters (Fig. 2) and gradually leveled off thereafter, indicating diminishing returns with additional clusters. consistent with the “elbow” criterion. The silhouette plot (Fig. 3) revealed that the two-cluster solution achieved the highest average silhouette width, suggesting strong separation between clusters and internal cohesion among cluster members. Based on the convergence of evidence from these criteria, alongside theoretical considerations and interpretability of the resulting clusters, we selected a two-cluster solution as the optimal model.

Two clusters were identified through *k*-means cluster analysis: Cluster 1–*Motivated but Unconfident* ( $n = 160$ ; 59.48%) and Cluster 2–*Confident but Disengaged* ( $n = 109$ ; 40.52%). Throughout the manuscript, including both *k*-means and LPA, these labels are used to describe students’ broader affective–motivational groupings. Specifically, “confidence” is conceptualized to encompass not only ability-based confidence (*i.e.*, self-efficacy), but also affective and identity-related orientations toward chemistry (*i.e.*, chemistry self-concept and attitudes). The two-cluster solution accounted for 22.6% of the total variance in the data as indicated by the ratio of between-cluster sum of squares to the total sum of squares. The within-cluster sum of squares for Clusters 1 and 2 were 559.71 and 477.73, respectively, suggesting acceptable internal homogeneity for both clusters (Kaufman and Rousseeuw, 1990).

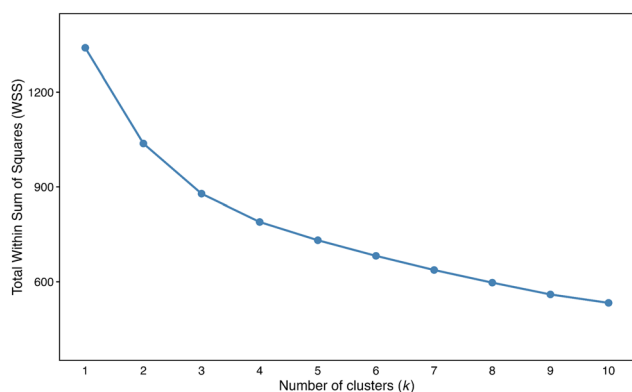


Fig. 2 Within-cluster Sum of Squares (WSS) to determine optimal number of clusters.

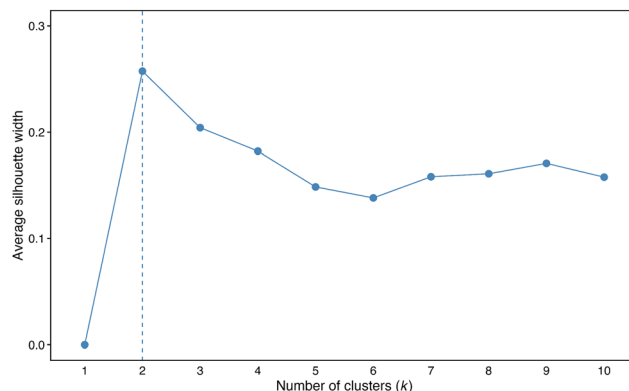


Fig. 3 Silhouette plot to determine optimal number of clusters.

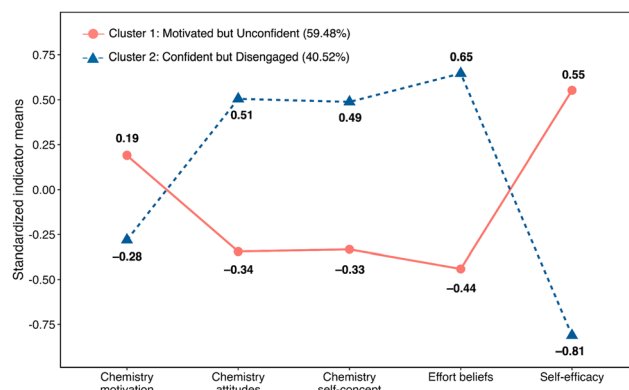


Fig. 4 Plot of indicator response means: two-cluster solution. Percentages indicate cluster proportions.

Response patterns can be visualized for each cluster. Fig. 4 presents a dot-line visualization for each cluster based on the mean scores for students’ affective indicators. Cluster 1, *Motivated but Unconfident*, was characterized by above-average chemistry motivation ( $z = 0.19$ ) and self-efficacy ( $z = 0.55$ ), but below-average scores in attitudes toward chemistry ( $z = -0.34$ ), effort beliefs ( $z = -0.44$ ), and chemistry self-concept ( $z = -0.33$ ). These patterns suggest that students in Cluster 1 are behaviorally engaged and motivated yet report low affective and self-evaluative confidence in chemistry. These patterns suggest that students in Cluster 1 are behaviorally engaged and motivated, and perceive themselves as capable of learning chemistry, yet hold a weaker affective and identity-related orientations toward the domain.

On the other hand, the second cluster, *Confident but Disengaged*, was characterized by elevated scores in chemistry attitudes ( $z = 0.51$ ), effort beliefs ( $z = 0.65$ ), and chemistry self-concept ( $z = 0.49$ ). These students reported positive feeling toward chemistry and emotional stability yet lacked motivation ( $z = -0.28$ ) and confidence to succeed (self-efficacy:  $z = -0.81$ ). Consistent with the broader definition of confidence used in this manuscript, this profile reflects affective and identity-based confidence in the domain despite reduced motivation and lower task-specific self-efficacy, suggesting disengagement from active learning behaviors.



**Covariate and distal outcome analysis.** We first investigated the relationship between the covariate (grading scheme: traditional vs. specifications grading), and students' affective outcomes. There were no significant differences in the proportion of students across the two clusters:  $t(267) = -0.70, p = 0.48$ . To examine whether the identified clusters differed meaningfully on the distal outcome (*i.e.*, ACS exam score), we compared students' exam performance across the two affective clusters. An independent samples *t*-test indicates that students in the *Motivated but Unconfident* cluster ( $M = 69.96, SD = 13.38$ ) scored statistically significantly higher than students in the *Confident but Disengaged* cluster ( $M = 62.13, SD = 13.74$ ),  $t(267) = 4.66, p < 0.001$ . These findings suggest that students' affective grouping is meaningfully associated with their performance on a standardized chemistry exam.

### Latent profile analysis

Latent profile enumeration was performed using all four model specifications to estimate the combination of item means, item variances, and item covariances across profiles. To determine the best-fitting solution within each model, we evaluated multiple fit indices as described in the Methods.

For Model 1 (diagonal, class invariant; see Table 2), there was no agreement on the best solution based on fit criteria. A two-profile solution was indicated by CAIC and VLMR, a three-profile solution was suggested by BIC, and a six-profile solution was identified by aBIC, BLRT, and BF. For Model 2 (diagonal, class varying), fit indices supported both a two-profile

model (BIC, CAIC, VLMR, BF) and a six-profile model (aBIC and BLRT). For Model 3 (non-diagonal, class invariant), a four-profile model was best supported by BLRT and BF. For Model 4 (non-diagonal, class varying), fit indices indicated a two-profile solution (BIC, BLRT and BF) and a four-profile solution (aBIC, VLMR, BF). These results reflect the complexity of model selection in LPA and underscore the importance of evaluating multiple model specifications and fit indices to identify the most interpretable and statistically sound solution.

From the set of candidate models, we conducted an additional round of comparison based on fit indices, model parsimony, and profile proportions (see Table 3). According to the fit indices, Model 4 (non-diagonal, class varying) with two profiles demonstrated the best overall fit, yielding the lowest absolute values for BIC (BIC = 3792.05) and CAIC (CAIC = 3833.05), and the highest cmP (cmP = 0.991). However, Model 2 (diagonal, class varying) with two profiles presented fit indices that were only slightly less favorable (BIC = 3831.49, CAIC = 3852.49, and cmP = 0.990), while requiring substantially fewer parameters (*i.e.*, 65 vs. 83 parameters, respectively). Overall, we settled upon Model 2 with two profiles because it offers a significantly more parsimonious solution with comparable statistical support.

Moreover, Model 4 with two profiles included a profile with a notably small proportion of the sample (2.22%), raising concerns about interpretative and model stability. Extremely small classes may reflect overfitting or statistical artifacts rather than meaningful subgroup differences, and can complicate interpretation and generalizability (Masyn, 2013; Nylund-Gibson and Choi, 2018).

Table 2 Summary of model fit for all LPA models

Model	K	Par	LL	BIC	aBIC	CAIC	AWE	BLRT	VLMR	BF	cmP
1	1	10	-1905.97	3867.88	3836.18	3877.88	<b>3953.83</b>	—	—	0.0	<0.001
	2	16	-1865.60	3820.72	3769.99	<b>3836.72</b>	3958.24	<0.001	<b>0.005</b>	0.2	0.062
	3	22	-1846.97	3817.02	3747.27	3839.02	4006.11	<0.001	0.192	0.7	0.396
	4	28	-1829.88	<b>3816.40</b>	3727.63	3844.40	4057.06	<0.001	0.193	>100	<b>0.540</b>
	5	34	-1819.06	3828.35	3720.55	3862.35	4120.57	<b>&lt;0.001</b>	0.169	>100	0.001
	6	40	-1809.36	3842.50	<b>3715.68</b>	3882.50	4186.29	0.066	0.039	—	—
2	1	10	-1905.97	3867.88	3836.18	3877.88	<b>3953.83</b>	—	—	0.0	<0.001
	2	21	-1857.00	<b>3831.49</b>	3764.91	<b>3852.49</b>	4011.98	<0.001	<b>0.032</b>	>100	<b>0.990</b>
	3	32	-1830.84	3840.71	3739.25	3872.71	4115.74	<b>&lt;0.001</b>	0.084	>100	0.010
	4	43	-1810.82	3862.20	3725.87	3905.20	4231.78	0.076	0.350	>100	<0.001
	5	54	-1793.12	3888.35	3717.14	3942.35	4352.47	0.235	0.166	>100	<0.001
	6	65	-1776.31	3916.28	<b>3710.19</b>	3981.28	4474.94	0.235	0.259	—	—
3	1	20	-1842.40	3796.70	3733.29	<b>3816.70</b>	<b>3968.60</b>	—	—	0.9	0.470
	2	26	-1825.51	<b>3796.48</b>	3714.04	3822.48	4019.94	<b>&lt;0.001</b>	<b>&lt;0.001</b>	>100	<b>0.526</b>
	3	32	-1813.59	3806.22	3704.76	3838.22	4081.25	0.667	0.283	>100	0.004
	4	38	-1802.70	3817.99	3697.51	3855.99	4144.59	1.000	0.081	>100	<0.001
	5	44	-1792.59	3831.34	3691.83	3875.34	4209.50	0.375	0.432	>100	<0.001
	6	50	-1784.32	3848.37	<b>3689.83</b>	3898.37	4278.10	1.000	0.469	—	—
4	1	20	-1842.40	3796.70	3733.29	<b>3816.70</b>	<b>3968.60</b>	—	—	0.1	0.089
	2	41	-1781.33	<b>3792.05</b>	3662.05	3833.05	4144.43	<b>&lt;0.001</b>	0.008	>100	<b>0.911</b>
	3	62	-1752.41	3851.70	3655.12	3913.70	4384.57	0.140	0.227	>100	<0.001
	4	83	-1724.50	3913.37	<b>3650.20</b>	3996.37	4626.73	<b>&lt;0.001</b>	<b>&lt;0.001</b>	>100	<0.001
	5	104	-1703.50	3988.85	3659.11	4092.85	4882.70	0.130	0.828	>100	<0.001
	6	125	-1681.00	4061.33	3665.00	4186.33	5135.67	1.000	0.240	—	—

Note. *K* = number of profiles; Par = number of 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 likelihood ratio test *p*-value; BF = Bayes Factor; cmP = correct model probability; Bold = best-fit statistic. Model 1 = restrictive class-invariant diagonal model with equal variances across latent classes and covariances fixed at zero; Model 2 = class-varying diagonal model with freely estimated variances and covariances fixed at zero; Model 3 = a class-invariant unrestricted model with equal variances and covariances; Model 4 = a class-varying unrestricted model with freely estimated variances and covariances.



Table 3 Summary of model fit for selected LPA models

Model	K	Par	BIC	aBIC	CAIC	AWE	BLRT	VLMR	BF	cmP	Smallest profile
1	2	16	3820.72	3769.99	3836.72	3958.24	<0.001	0.005	0.2	0.062	41.92%
	3	22	3817.02	3747.27	3839.02	4006.11	<0.001	0.192	0.7	0.396	2.59%
	6	40	3842.50	3715.68	3882.50	4186.29	0.066	0.039	—	<0.001	0.37%
2	2	21	3831.49	3764.91	3852.49	4011.98	<0.001	0.032	>100	0.990	49.04%
	6	65	3916.28	3710.19	3981.28	4474.94	0.235	0.259	—	<0.001	1.82%
3	4	38	3817.99	3697.51	3855.99	4144.59	1.000	0.081	>100	<0.001	0.37%
4	2	41	3792.05	3662.05	3833.05	4144.43	<0.001	0.008	>100	0.911	2.22%
	4	83	3913.37	3650.20	3996.37	4626.73	<0.001	<0.001	>100	<0.001	2.22%

Considering statistical fit, parsimony, theoretical alignment, and interpretability, we selected the Model 2 two-profile solution as the final model for subsequent analyses.

The selected LPA model identified the following two profiles that closely resemble the clusters found in the *k*-means cluster analysis: (1) *Motivated but Unconfident* ( $n = 138$ ; 50.96%) and (2) *Confident but Disengaged* ( $n = 131$ ; 49.04%), using the same label definitions introduced in the *k*-means analysis. Fig. 5 presents a dot-line visualization of each profile based on the mean scores for students' affective outcomes. Students in the *Motivated but Unconfident* profile demonstrated high self-efficacy ( $z = 0.66$ ) and moderately positive motivation ( $z = 0.23$ ) but reported lower scores in attitudes toward chemistry ( $z = -0.32$ ), effort beliefs ( $z = -0.46$ ), and chemistry self-concept ( $z = -0.25$ ). This pattern suggests that these students reported some belief in their ability to succeed in chemistry, while also expressing weaker chemistry self-concept and less positive attitudes and effort beliefs. Thus, the profile appears characterized by a mixed affective pattern than uniformly high confidence or engagement. Here, "Unconfident" refers primarily to weaker chemistry self-concept rather than low self-efficacy alone.

Students in the *Confident but Disengaged* profile exhibited positive attitude towards chemistry ( $z = 0.33$ ), strong effort beliefs ( $z = 0.48$ ), and high chemistry self-concept ( $z = 0.26$ ), but reported notably low general self-efficacy ( $z = -0.69$ ) and low motivation to learn chemistry ( $z = -0.24$ ). This pattern suggests that while these students hold positive affective and identity-based orientations toward chemistry, they report lower self-efficacy and lower motivation to learn chemistry, which

may contribute to disengagement from active learning behaviors.

**Covariate and distal outcome analysis.** We evaluated the relation of the covariate (grading scheme: specifications vs. traditional) and the distal outcome (ACS exam score) using the automated three-step approach (Fig. 6). First, we tested whether the grading scheme predicted latent profile membership. No statistically significant relationship was found ( $\beta = 0.03$ ,  $p = 0.91$ ).

Next, we examined whether profile membership was associated with the ACS exam. An omnibus *F*-test, the linear model equivalent of a Wald test, revealed a significant difference in exam scores across profiles:  $F(1, 267) = 25.60$ ,  $p < 0.001$ . Specifically, students in the *Motivated but Unconfident* profile scored significantly higher ( $M = 70.82$ ,  $SD = 13.29$ ) on the ACS exam than those in the *Confident but Disengaged* profile ( $M = 62.53$ ,  $SD = 13.58$ ), suggesting that affective profile membership is meaningfully related to chemistry learning outcomes.

### Comparing *k*-means cluster analysis and LPA groupings

Both LPA and *k*-means clustering identified two distinct subgroups of students based on affective indicators. These findings demonstrate strong alignment across the two different methodological approaches for this specific example in this context. To compare the groupings, we used two evaluated three key dimensions: (1) consistency of group membership, (2) similarity of group characteristics, and (3) impact on distal outcomes.

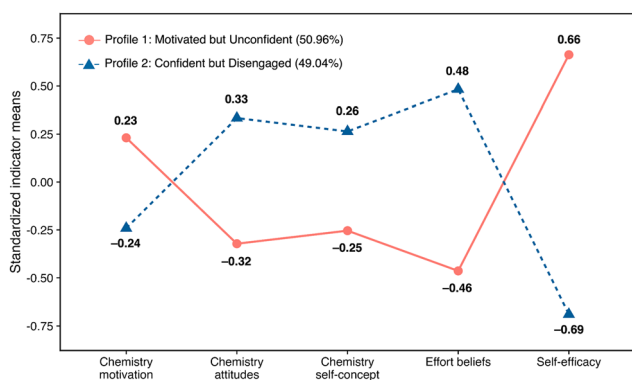


Fig. 5 Plot of conditional item response means of the two profile LPA model. Percentages indicate profile proportions based on the diagonal, class varying model with two profiles.

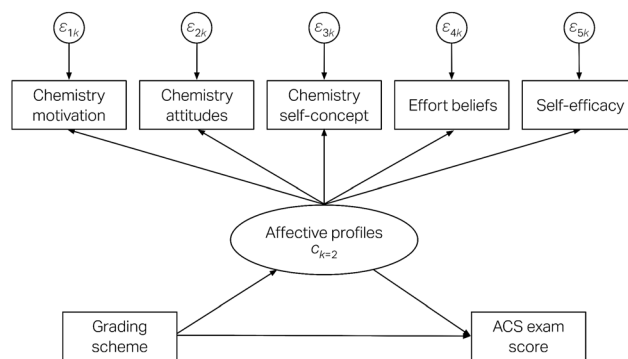


Fig. 6 Final path diagram for the diagonal, class varying LPA model for the latent grouping variable (*i.e.*, affective profiles) with five indicators (*i.e.*, chemistry motivation, chemistry attitudes, chemistry self-concept, effort beliefs, and self-efficacy), one covariate (*i.e.*, grading scheme), and one distal outcome (*i.e.*, ACS exam score).



Table 4 Comparison of *k*-means clustering and LPA results

	LPA Profile 1: <i>Motivated but Unconfident</i>	LPA Profile 2: <i>Confident but Disengaged</i>	Total
<i>k</i> -Means Cluster 1: <i>Motivated but Unconfident</i>	137 (99.3%)	23 (17.6%)	160
<i>k</i> -Means Cluster 2: <i>Confident but Disengaged</i>	1 (0.7%)	108 (82.4%)	109
Total	138	131	269

**Group membership consistency.** To assess group consistency, we examined whether individuals were assigned to the same group across both methods (Table 4). Results showed that 99.3% of individuals in *k*-means Cluster 1 were also classified into LPA Profile 1, and 82.4% of those in *k*-means Cluster 2 aligned with LPA Profile 2. However, some discrepancies were observed; for instance, 17.6% of individuals in LPA Profile 2 were assigned to *k*-means Cluster 1, indicating subtle differences in classification.

These discrepant cases primarily involved students whose response patterns fell between the two profiles rather than closely matching the mean pattern of either group. *k*-Means clustering assigns individuals deterministically based on distance to cluster centroids, requiring each individual to be placed into a single cluster regardless of classification ambiguity. In contrast, LPA estimates posterior probabilities of profile membership, and thus explicitly represents uncertainty in classification. For individuals whose responses fall near the classification boundary, LPA estimates the probability of membership in each profile, thereby capturing classification uncertainty in a way that distance-based methods do not. In this study, we used the highest probability assignment to determine profile membership; however, these uncertainty estimates could be incorporated in future work for evaluating the strength of profile membership or when examining relationships with distal outcomes or other auxiliary variables.

From a modeling perspective, these discrepancies likely stem from the distinct assumptions underlying each method. *k*-Means clustering assumes equal variance across all dimensions and does not account for covariance among variables. In contrast, the selected LPA model (Model 2: diagonal, class varying) allows for freely estimated indicator variances across profiles while fixing covariances at zero. This added flexibility enables LPA to weigh indicator variables differently depending on their within-profile variability, allowing the technique to capture more nuanced subgroup structures.

**Group membership similarity.** A comparison of affective grouping characteristics (see Fig. 4 for *k*-means clustering and Fig. 5 for LPA) revealed highly similar *z*-score patterns across the key variables. This similarity suggests that both methods identified comparable underlying structures and further support the convergence between *k*-means cluster analysis and LPA for this specific example.

**Distal outcome impact.** We also compared the impact of group membership on the distal outcome, ACS exam performance. Both methods revealed a consistent pattern: students in the *Motivated but Unconfident* group scored significantly higher on the exam than those in the *Confident but Disengaged* group. The mean exam scores were highly similar across methods

(*k*-means cluster analysis: 69.96 vs. 62.12; LPA: 70.82 vs. 62.53), and both differences were statistically significant. These findings reinforce the validity of the identified groupings and their relevance to academic performance.

### Methodological considerations

This comparison highlights both the alignment and meaningful differences between *k*-means cluster analysis and LPA. *k*-Means clustering offers a simple, distance-based approach that assumes equal variances and spherical clusters, while LPA provides greater flexibility by allowing variance to differ across profiles. The selected LPA model (Model 2: diagonal, class varying) freely estimates the within-profile variances, enabling it to capture more complex and nuanced patterns in the data that may be overlooked by *k*-means clustering.

LPA is model-based and uses fit indices and likelihood tests to evaluate competing models. Conversely, *k*-means clustering is algorithmic, relying on distance metrics and heuristic criteria (e.g., elbow method) for model selection. Together, these findings underscore the importance of considering both statistical assumptions and theoretical interpretability when selecting and comparing variable-centered and person-centered analytical grouping techniques.

Both methods allow multiple random starting values, but in *k*-means cluster analysis these starts are more likely to produce different hard partitions of the data due to the algorithm's sensitivity to initialization. In contrast, LPA relies on likelihood-based estimation, and convergence to the best log-likelihood solution across many random starts is often taken as evidence that global or near-global maximum has been identified. However, it does not guarantee stability. As noted in the mixture modeling literature, different sets of starting values can yield alternative local maxima that may correspond to substantively different profile solutions (e.g., McLachlan and Peel, 2000; Biernacki *et al.*, 2003). Accordingly, researchers are encouraged to use a large number of random starts (at least 500 initial starts and 100 starts in the final stage) and to examine whether competing solutions may provide alternative interpretations of the data.

### Software limitations

While using open-access statistical packages, such as R, has its advantages, it is also important to recognize its limitations. For example, the *tidyLPA* package in R does not adjust for classification error and assumes perfect classification when estimating relationships between profiles and auxiliary variables (e.g., covariates and distal outcomes). It also requires manual steps (i.e., no automatic function to perform the task), such as conducting the latent class regression or ANOVA separate, to



analyze these relationships. In contrast, Mplus is a paid commercial statistical package that natively supports the three-step approach and offers advanced capabilities such as multiple-group mixture modeling and measurement invariance testing across profiles.

### Study limitations

While this study offers valuable insights into the comparison between *k*-means cluster analysis and LPA in chemistry education research, several limitations should be acknowledged. First, findings are based on data from a single institution and course context, which may limit generalizability to other educational contexts or disciplines. Additionally, this study involves only one specific implementation of specifications grading and traditional grading; therefore, findings may not extend to other versions of these specific grading approaches or to different instructional contexts. Second, the comparison involved one *k*-means solution and one selected LPA solution based on the present data and modeling decisions. Different preprocessing choices, number of groups, model specifications, or indicator sets could yield different subgroup solutions. Thus, the findings should not be interpreted as establishing the superiority of either method in general, but rather as illustrating how these two approaches performed in this particular application. Third, although both methods identified broadly similar groupings, interpretation of those groupings remains sensitive to labeling decisions and to the specific affective indicators included in the analysis. More research is needed to examine how robust these subgrouping patterns are across contexts, measures, and alternative analytic modeling specifications.

Second, the *tidyLPA* package in R assumes perfect classification and does not adjust for classification error when relating profiles to auxiliary variables. This may introduce bias in estimating associations with covariates and distal outcomes. Software such as Mplus offers built-in functionality to overcome these limitations and may provide more robust estimates. Third, this study relied on self-reported affective measures for five constructs. The inclusion or exclusion of additional affective indicators may influence the resulting groupings thus changes the interpretation of subgroup characteristics. Finally, while both grouping techniques produced similar results, the interpretation of clusters and profiles is inherently subjective and influenced by theoretical framing. Future research should consider replicating these analyses across diverse samples, using alternative statistical software (*e.g.*, Mplus) that accounts for classification errors.

## LPA for chemistry education research

### Implications for researchers

Using a worked example, this study compares *k*-means cluster analysis with LPA on a data set relevant to chemistry education. While both can be valuable tools it should be noted that LPA is part of the larger modeling framework of mixture modeling. Other mixture models include latent class analysis (LCA) for

use with categorical data (Nylund-Gibson and Choi, 2018), and latent transition analysis (LTA), which extends mixture modeling to longitudinal data (Nylund-Gibson *et al.*, 2023; Odeleye *et al.*, 2025).

Our goal in this article is to provide researchers with a glimpse into the strengths and limitations of various grouping approaches, along with the statistical assumptions and theoretical considerations that can inform all steps of the research process. Although this work is not a step-by-step primer on implementing LPA, numerous resources are available to support researchers in learning more. For example, Nylund-Gibson and colleagues provide several examples and code for model estimation for LCA (Nylund-Gibson and Choi, 2018) and LTA (Nylund-Gibson *et al.*, 2023) and compared *k*-modes clustering with LCA (Wang *et al.*, 2025a). Additionally, the IMMERSE (Institute of Mixture Modeling for Equity-Oriented Researchers, Scholars, and Educators) training program has a large set of free online resources available (IMMERSE, 2025). The mixture modeling framework enables researchers to adopt person-centered grouping approaches rather than grouping variables which may lead to more equity-centered quantitative analyses (Slominski *et al.*, 2024) while also embedding the mixture model into a larger modeling framework similar to structural equation modeling (see Arch *et al.*, 2025). We encourage chemistry education researchers to consider LPA and other mixture models in their own work, guided by their research questions and available data. To support reproducibility, we provide the code used to generate the *k*-means clustering and LPA results (Zhang and Yik, 2026).

### Implications for practitioners

This study underscores the importance of bridging chemistry education research and practice through collaboration (James *et al.*, 2024; Popova, 2024). For instance, this collaboration began when practitioner/author JDH approached researcher/author BJY with questions about affective outcomes related to specifications grading. This inquiry led to a partnership that not only addressed practitioner concerns but also introduced new quantitative methods to the field from a research standpoint. As a result, practitioners JDH and EDN had their questions explored through collaboration with researcher BJY, who then partnered with researchers YZ, KNG, and MI to apply and learn new methodologies. We advocate for these collaborations and partnerships between researchers and practitioners to become commonplace in chemistry education.

Instructors interested in grouping students based on multiple measures can consider approaches such as *k*-means cluster analysis and LPA. These methods help reduce dimensionality; for example, in a class of more than 100 students, grouping can help simplify the challenge in understanding the heterogeneity of the classroom and tailoring interventions to meet individual needs. By identifying a few meaningful student groups, instructors can more effectively target support or interventions.

While LPA and mixture modeling may require statistical expertise and collaboration with researchers, *k*-means clustering is more accessible and may still provide valuable insights.



This study demonstrates strong alignment between the groupings produced by both methods in this context, suggesting that even simpler approaches can yield useful results for classroom decision-making.

## Author contributions

JDH and BJY conceived the broader project. BJY, MI, and KNG conceived the specific study. JDH, BJY, and EDN collected the data. BJY and YZ cleaned the data. YZ, KNG, MI, and BJY analyzed the data. BJY and YZ wrote the manuscript with contributions from LK. All authors read, edited, and approved the final manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

Raw data that support the findings of this study are not publicly available due to participants' confidentiality and restrictions set by the participating institution.

Supplementary information (SI) includes syllabi for the traditionally- and specifications-graded courses and online survey. See DOI: <https://doi.org/10.1039/d5rp00432b>.

## Acknowledgements

We would like to thank all the students who participated in this study, and Victoria Alvarado who assisted with the literature review. This material is based upon work supported by the National Science Foundation under Grant No. 2224786 (KNG and MI) and 2339405 (EDN). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. JDH thanks the Office of the President at The Pennsylvania State University for funding through the Opportunity Grant Professional Development Program.

## References

- An J., Guzman-Joyce G., Brooks A., To K., Vu L. and Luxford C. J., (2022), Cluster analysis of learning approaches and course achievement of general chemistry students at a Hispanic serving institution, *J. Chem. Educ.*, **99**(2), 669–677, DOI: [10.1021/acs.jchemed.1c00759](https://doi.org/10.1021/acs.jchemed.1c00759).
- Arch D. A. N., Nylund-Gibson K. and Ing M., (2025), Moderation with a latent class variable: a tutorial and example, *Behav. Res. Methods*, **58**, 108, DOI: [10.3758/s13428-025-02886-x](https://doi.org/10.3758/s13428-025-02886-x).
- Asparouhov T. and Muthén B., (2014), Auxiliary variables in mixture modeling: three-step approaches using Mplus, *Struct. Equ. Model.*, **21**(3), 329–341, DOI: [10.1080/10705511.2014.915181](https://doi.org/10.1080/10705511.2014.915181).
- Auf der Heyde T. P. E., (1990), Analyzing chemical data in more than two dimensions: a tutorial on factor and cluster analysis, *J. Chem. Educ.*, **67**(6), 461–469, DOI: [10.1021/ed067p461](https://doi.org/10.1021/ed067p461).
- Bandura A., (1977), Self-efficacy: toward a unifying theory of behavioral change, *Psych. Rev.*, **84**(2), 191–215, DOI: [10.1037/0033-295X.84.2.191](https://doi.org/10.1037/0033-295X.84.2.191).
- Bauer C. F., (2008), Attitude toward chemistry: a semantic differential instrument for assessing curriculum impacts, *J. Chem. Educ.*, **85**(10), 1440–1445, DOI: [10.1021/ed085p1440](https://doi.org/10.1021/ed085p1440).
- Biernacki C., Celeux G. and Govaert G., (2003), Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models, *Comput. Stat. Data Anal.*, **41**(3–4), 561–575, DOI: [10.1016/S0167-9473\(02\)00163-9](https://doi.org/10.1016/S0167-9473(02)00163-9).
- Blackwell L. S., (2002), Psychological mediators of student achievement during the transition to junior high school: The role of implicit theories, Unpublished Doctoral Dissertation, Columbia University.
- Blackwell L. S., Trzesniewski K. H. and Dweck C. S., (2007), Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention, *Child Dev.*, **78**(1), 246–263, DOI: [10.1111/j.1467-8624.2007.00995.x](https://doi.org/10.1111/j.1467-8624.2007.00995.x).
- Blashfield R. K. and Aldenderfer M. S., (1988), *The methods and problems of cluster analysis*, in Nesselroade J. R. and Cattell R. B. (ed.), *Handbook of multivariate experimental psychology*, Boston, MA: Springer, pp. 447–473.
- Blum S. D., (2020), *Ungrading: Why rating students undermines learning (and what to do instead)*, Morgantown, WV: West Virginia University Press.
- Boesdorfer S. B., Baldwin E. and Lieberum K. A., (2018), Emphasizing learning: using standards-based grading in a large nonmajors' general chemistry survey course, *J. Chem. Educ.*, **95**(8), 1291–1300, DOI: [10.1021/acs.jchemed.8b00251](https://doi.org/10.1021/acs.jchemed.8b00251).
- Bong M. and Skaalvik E. M., (2003), Academic self-concept and self-efficacy: How different are they really? *Educ. Psychol. Rev.* **15**(1), 1–40, DOI: [10.1023/A:1021302408382](https://doi.org/10.1023/A:1021302408382).
- Brandriet A. R. and Bretz S. L., (2014), Measuring meta-ignorance through the lens of confidence: examining students' redox misconceptions about oxidation numbers, charge, and electron transfer, *Chem. Educ. Res. Pract.*, **15**(4), 729–746, DOI: [10.1039/C4RP00129J](https://doi.org/10.1039/C4RP00129J).
- Brandriet A. R., Xu X., Bretz S. L. and Lewis J. E., (2011), Diagnosing changes in attitude in first-year college chemistry students with a shortened version of Bauer's semantic differential, *Chem. Educ. Res. Pract.*, **12**(2), 271–278, DOI: [10.1039/C1RP90032C](https://doi.org/10.1039/C1RP90032C).
- Braun I., Lewis S. E. and Graulich N., (2025), A question of pattern recognition: investigating the impact of structure variation on students' proficiency in deciding about resonance stabilization, *Chem. Educ. Res. Pract.*, **26**(1), 158–182, DOI: [10.1039/D4RP00155A](https://doi.org/10.1039/D4RP00155A).
- Burnham K. P., Anderson D. R. and Huyvaert K. P., (2011), AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons,



- Behav. Ecol. Sociobiol.*, **65**(1), 23–35, DOI: [10.1007/s00265-010-1029-6](https://doi.org/10.1007/s00265-010-1029-6).
- Chan J. Y. K. and Bauer C. F., (2014), Identifying at-risk students in general chemistry via cluster analysis of affective characteristics, *J. Chem. Educ.*, **91**(9), 1417–1425, DOI: [10.1021/ed500170x](https://doi.org/10.1021/ed500170x).
- Charrad M., Ghazzali N., Boiteau V. and Niknafs A., (2014), NbClust: an R package for determining the relevant number of clusters in a data set, *J. Stat. Softw.*, **61**(6), 1–36, DOI: [10.18637/jss.v061.i06](https://doi.org/10.18637/jss.v061.i06).
- Clark D. and Talbert R., (2023), *Grading for growth: A guide to alternative grading practices that promote authentic learning and student engagement in higher education*, New York, NY: Stylus.
- Closser K. D., Hawker M. J. and Muchalski H., (2024), Quantized grading: an ab initio approach to using specifications grading in physical chemistry, *J. Chem. Educ.*, **101**(2), 474–482, DOI: [10.1021/acs.jchemed.3c00872](https://doi.org/10.1021/acs.jchemed.3c00872).
- Collins L. M. and Lanza S. T., (2010), *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*, Hoboken, NJ: Wiley.
- Connor M. C., Glass B. H. and Shultz G. V., (2021), Development of the NMR Lexical Representational Competence (NMR-LRC) instrument as a formative assessment of lexical ability in  $^1\text{H}$  NMR spectroscopy, *J. Chem. Educ.*, **98**(9), 2786–2798, DOI: [10.1021/acs.jchemed.1c00332](https://doi.org/10.1021/acs.jchemed.1c00332).
- Dawes J., (2008), Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales, *Int. J. Mark. Res.*, **50**(1), 61–104, DOI: [10.1177/147078530805000106](https://doi.org/10.1177/147078530805000106).
- Deci E. L. and Ryan R. M., (1985), *Intrinsic motivation and self-determination in human behavior*, New York, NY: Plenum.
- Deci E. L. and Ryan R. M., (2000), The “what” and “why” of goal pursuits: human needs and the self-determination of behavior, *Psychol. Inq.*, **11**(4), 227–268, DOI: [10.1207/S15327965PLI1104\\_01](https://doi.org/10.1207/S15327965PLI1104_01).
- Diegelman-Parente A., (2011), The use of mastery learning with competency-based grading in an organic chemistry course, *J. Coll. Sci. Teach.*, **40**(5), 50–58.
- Drechsler M. and Van Driel J., (2009), Teachers' perceptions of the teaching of acids and bases in Swedish upper secondary schools, *Chem. Educ. Res. Pract.*, **10**(2), 86–96, DOI: [10.1039/B908246H](https://doi.org/10.1039/B908246H).
- Eccles J. S. and Wigfield A., (2002), Motivational beliefs, values, and goals, *Annu. Rev. Psychol.*, **53**(1), 109–132, DOI: [10.1146/annurev.psych.53.100901.135153](https://doi.org/10.1146/annurev.psych.53.100901.135153).
- Everitt B. S., Landau S., Leese M. and Stahl D., (2011), *Cluster analysis*, Chichester, UK: Wiley.
- Feldman J., (2019a), Beyond standards-based grading: Why equity must be a part of grading reform, *Phi Delta Kappan*, **100**(8), 52–55, DOI: [10.1177/0031721719846890](https://doi.org/10.1177/0031721719846890).
- Feldman J., (2019b), What traditional grading gets wrong, *Educ. Week*, **38**(19), 18–19.
- Ferreira J. E. V. and Lawrie G. A., (2019), Profiling the combinations of multiple representations used in large-class teaching: pathways to inclusive practices, *Chem. Educ. Res. Pract.*, **20**(4), 902–923, DOI: [10.1039/C9RP00001](https://doi.org/10.1039/C9RP00001).
- Ferrell B. and Barbera J., (2015), Analysis of students' self-efficacy, interest, and effort beliefs in general chemistry, *Chem. Educ. Res. Pract.*, **16**(2), 318–337, DOI: [10.1039/C4RP00152D](https://doi.org/10.1039/C4RP00152D).
- Gaias L. M., Lindstrom Johnson S., Bottiani J. H., Debnam K. J. and Bradshaw C. P., (2019), Examining teachers' classroom management profiles: incorporating a focus on culturally responsive practice, *J. Sch. Psychol.*, **76**, 124–139, DOI: [10.1016/j.jsp.2019.07.017](https://doi.org/10.1016/j.jsp.2019.07.017).
- Galloway K. R. and Bretz S. L., (2015a), Measuring meaningful learning in the undergraduate general chemistry and organic chemistry laboratories: a longitudinal study, *J. Chem. Educ.*, **92**(12), 2019–2030, DOI: [10.1021/acs.jchemed.5b00754](https://doi.org/10.1021/acs.jchemed.5b00754).
- Galloway K. R. and Bretz S. L., (2015b), Using cluster analysis to characterize meaningful learning in a first-year university chemistry laboratory course, *Chem. Educ. Res. Pract.*, **16**(4), 879–892, DOI: [10.1039/C5RP00077G](https://doi.org/10.1039/C5RP00077G).
- Gibbons R. E., Villafañe S. M., Stains M., Murphy K. L. and Raker J. R., (2018), Beliefs about learning and enacted instructional practices: an investigation in postsecondary chemistry education, *J. Res. Sci. Teach.*, **55**(8), 1111–1133, DOI: [10.1002/tea.21444](https://doi.org/10.1002/tea.21444).
- Gibbons R. E., Reed J. J., Srinivasan S., Murphy K. L. and Raker J. R., (2022), Assessment tools in context: results from a national survey of postsecondary chemistry faculty, *J. Chem. Educ.*, **99**(8), 2843–2852, DOI: [10.1021/acs.jchemed.2c00269](https://doi.org/10.1021/acs.jchemed.2c00269).
- Guerris M., Cuadros J., González-Sabaté L. and Serrano V., (2020), Describing the public perception of chemistry on twitter, *Chem. Educ. Res. Pract.*, **21**(3), 989–999, DOI: [10.1039/C9RP00282K](https://doi.org/10.1039/C9RP00282K).
- Gulacar O., Vernoy B., Tran E., Wu A., Huie E. Z., Santos E. V., Wadhwa A., Sathe R. and Milkey A., (2022), Investigating differences in experts' chemistry knowledge structures and comparing them to those of general chemistry students, *J. Chem. Educ.*, **99**(8), 2950–2963, DOI: [10.1021/acs.jchemed.2c00251](https://doi.org/10.1021/acs.jchemed.2c00251).
- Guo Y., O'Halloran K. P., Eaker R. M., Anfuso C. L., Kirberger M. and Gluick T., (2022), Affective elements of the student experience that contribute to withdrawal rates in the general chemistry sequence: a multimethod study, *J. Chem. Educ.*, **99**(6), 2217–2230, DOI: [10.1021/acs.jchemed.1c01227](https://doi.org/10.1021/acs.jchemed.1c01227).
- Hackerson E. L., Slominski T., Johnson N., Buncher J. B., Ismael S., Singelmann L., Leontyev A., Knopps A. G., McDarby A., Nguyen J. J., Condry D. L. J., Nyachwaya J. M., Wissman K. T., Falkner W., Grieger K., Montplaisir L., Hodgson A. and Momsen J. L., (2024), Alternative grading practices in undergraduate STEM education: a scoping review, *Discip. Interdiscip. Sci. Educ. Res.*, **6**(1), 15, DOI: [10.1186/s43031-024-00106-8](https://doi.org/10.1186/s43031-024-00106-8).
- Hensen C. and Barbera J., (2019), Assessing affective differences between a virtual general chemistry experiment and a similar hands-on experiment, *J. Chem. Educ.*, **96**(10), 2097–2108, DOI: [10.1021/acs.jchemed.9b00561](https://doi.org/10.1021/acs.jchemed.9b00561).
- Hickendorff M., Edelsbrunner P. A., McMullen J., Schneider M. and Trezise K., (2018), Informative tools for characterizing



- individual differences in learning: latent class, latent profile, and latent transition analysis, *Learn. Individ. Differ.*, **66**, 4–15, DOI: [10.1016/j.lindif.2017.11.001](https://doi.org/10.1016/j.lindif.2017.11.001).
- Howitz W. J., McKnelly K. J. and Link R. D., (2021), Developing and implementing a specifications grading system in an organic chemistry laboratory course, *J. Chem. Educ.*, **98**(2), 385–394, DOI: [10.1021/acs.jchemed.0c00450](https://doi.org/10.1021/acs.jchemed.0c00450).
- Hunter R. A., Pompano R. R. and Tuchler M. F., (2022), Alternative assessment of active learning, *Active learning in the analytical chemistry curriculum*, Washington, DC: American Chemical Society, vol. 1409, ch. 15, pp. 269–295, DOI: [10.1021/bk-2022-1409.ch015](https://doi.org/10.1021/bk-2022-1409.ch015).
- IMMERSE (Institute of Mixture Modeling for Equity-Oriented Researchers, Scholars, and Educators), (2025), IMMERSE online resources (IES No. 305B220021). Institute of Education Sciences. Available at: <https://mixture-modeling.netlify.app/> (Accessed 25 September 2025).
- Inoue A. B., (2022), *Labor-based grading contracts: Building equity and inclusion in the compassionate writing classroom*, Fort Collins, CO: The WAC Clearinghouse.
- Jaison J. A., Cruz K. A. and Liu Y., (2025), Investigating students' academic motivation, homework, and academic achievement in an online general chemistry II course, *J. Chem. Educ.*, **102**(2), 485–494, DOI: [10.1021/acs.jchemed.4c00736](https://doi.org/10.1021/acs.jchemed.4c00736).
- James N. M., McKenna M. S. and Mishra A., (2024), Toward collaborative dialogue: unpacking the researcher–educator divide to advance chemistry education, *J. Chem. Educ.*, **101**(8), 2960–2965, DOI: [10.1021/acs.jchemed.3c01321](https://doi.org/10.1021/acs.jchemed.3c01321).
- Jeffery K. A. and Bauer C. F., (2020), Students' responses to emergency remote online teaching reveal critical factors for all teaching, *J. Chem. Educ.*, **97**(9), 2472–2485, DOI: [10.1021/acs.jchemed.0c00736](https://doi.org/10.1021/acs.jchemed.0c00736).
- Jones B. D., Wilkins J. L. M., Long M. H. and Wang F., (2012), Testing a motivational model of achievement: How students' mathematical beliefs and interests are related to their achievement, *Eur. J. Psychol. Educ.*, **27**(1), 1–20, DOI: [10.1007/s10212-011-0062-9](https://doi.org/10.1007/s10212-011-0062-9).
- Jurišević M., Vrtačnik M., Kwiatkowski M. and Gros N., (2012), The interplay of students' motivational orientations, their chemistry achievements and their perception of learning within the hands-on approach to visible spectrometry, *Chem. Educ. Res. Pract.*, **13**(3), 237–247, DOI: [10.1039/C2RP20004J](https://doi.org/10.1039/C2RP20004J).
- Kaufman L. and Rousseeuw P. J., (1990), *Finding groups in data: An introduction to cluster analysis*, Hoboken, NJ: Wiley.
- Knox J., Lawson T. K., Goodwin A. B., Golden A. R., Arch D. A. N. and Fallon L., (2025), Supporting cultural responsiveness in the classroom: an exploratory study of teacher profiles, *J. Educ. Psychol. Cons.*, DOI: [10.1080/10474412.2024.2449340](https://doi.org/10.1080/10474412.2024.2449340).
- Lastusaari M. and Murtonen M., (2013), University chemistry students' learning approaches and willingness to change major, *Chem. Educ. Res. Pract.*, **14**(4), 496–506, DOI: [10.1039/C3RP00045A](https://doi.org/10.1039/C3RP00045A).
- Lee S. and Guo Y., (2024), Evolution of teaching independent undergraduate chemistry research courses and student skills developed through the COVID-19 pandemic and forward, *J. Chem. Educ.*, **101**(4), 1726–1734, DOI: [10.1021/acs.jchemed.3c00694](https://doi.org/10.1021/acs.jchemed.3c00694).
- Lewis S. E., (2018), Goal orientations of general chemistry students via the achievement goal framework, *Chem. Educ. Res. Pract.*, **19**(1), 199–212, DOI: [10.1039/C7RP00148G](https://doi.org/10.1039/C7RP00148G).
- Lewis S. E., Shaw J. L., Heitz J. O. and Webster G. H., (2009), Attitude counts: self-concept and success in general chemistry, *J. Chem. Educ.*, **86**(6), 744–749, DOI: [10.1021/ed086p744](https://doi.org/10.1021/ed086p744).
- Linenberger K. J. and Holme T. A., (2014), Results of a national survey of biochemistry instructors to determine the prevalence and types of representations used during instruction and assessment, *J. Chem. Educ.*, **91**(6), 800–806, DOI: [10.1021/ed400201v](https://doi.org/10.1021/ed400201v).
- Link L. J. and Guskey T. R., (2019), How traditional grading contribute to student inequities and how to fix it, *Curric. Context*, **45**(1), 12–19, [https://uknowledge.uky.edu/edp\\_facpub/53](https://uknowledge.uky.edu/edp_facpub/53).
- Liu Y., Ferrell B., Barbera J. and Lewis J. E., (2017), Development and evaluation of a chemistry-specific version of the academic motivation scale (AMS-Chemistry), *Chem. Educ. Res. Pract.*, **18**(1), 191–213, DOI: [10.1039/C6RP00200E](https://doi.org/10.1039/C6RP00200E).
- Liu Y., Raker J. R. and Lewis J. E., (2018), Evaluating student motivation in organic chemistry courses: moving from a lecture-based to a flipped approach with peer-led team learning, *Chem. Educ. Res. Pract.*, **19**(1), 251–264, DOI: [10.1039/C7RP00153C](https://doi.org/10.1039/C7RP00153C).
- Liu Y., Niu M. and Sun H., (2026), A pattern analysis of Chinese high school students' chemistry achievement goals based on latent profile analysis, *J. Chem. Educ.*, **103**(1), 89–100, DOI: [10.1021/acs.jchemed.5c00735](https://doi.org/10.1021/acs.jchemed.5c00735).
- MacQueen J., (1967), Some methods for classification and analysis of multivariate observations, *Berkeley Symp. Math. Statist. Prob.*, **5**(1), 281–297.
- Martin P. P., Kranz D., Wulff P. and Graulich N., (2024), Exploring new depths: applying machine learning for the analysis of student argumentation in chemistry, *J. Res. Sci. Teach.*, **61**(8), 1757–1792, DOI: [10.1002/tea.21903](https://doi.org/10.1002/tea.21903).
- Masyn K., (2013), *Latent class analysis and finite mixture modeling*, in Little T. (ed.), *The oxford handbook of quantitative methods in psychology*, Oxford, UK: Oxford University Press, vol. 2, pp. 551–611, DOI: [10.1093/oxfordhb/9780199934898.013.0025](https://doi.org/10.1093/oxfordhb/9780199934898.013.0025).
- Matz R. L., Koester B. P., Fiorini S., Grom G., Shepard L., Stangor C. G., Weiner B. and McKay T. A., (2017), Patterns of gendered performance differences in large introductory courses at five research universities, *AERA Open*, **3**(4), 1–12, DOI: [10.1177/2332858417743754](https://doi.org/10.1177/2332858417743754).
- McLachlan G. and Peel D., (2000), *Finite mixture models*, New York: Wiley, DOI: [10.1002/0471721182](https://doi.org/10.1002/0471721182).
- Mueller C. M. and Dweck C. S., (1998), Praise for intelligence can undermine children's motivation and performance, *J. Pers. Soc. Psychol.*, **75**(1), 33–52, DOI: [10.1037/0022-3514.75.1.33](https://doi.org/10.1037/0022-3514.75.1.33).
- Muthén L. K. and Muthén B., (2017), *Mplus user's guide: Statistical analysis with latent variables, user's guide*, Los Angeles, CA: Muthén & Muthén.



- Nielsen S. E. and Yeziński E. J., (2016), Beyond academic tracking: using cluster analysis and self-organizing maps to investigate secondary students' chemistry self-concept, *Chem. Educ. Res. Pract.*, **17**(4), 711–722, DOI: [10.1039/C6RP00058D](https://doi.org/10.1039/C6RP00058D).
- Nilson L. B., (2015), *Specifications grading: Restoring rigor, motivating students, and saving faculty time*, Sterling, VA: Stylus Publishing.
- Nilson L. B. and Packowski J. A., (2026), *Specifications grading 2.0: Restoring rigor, motivating students, saving faculty time, and developing career competencies*, New York, NY: Routledge.
- Noell S. L., Rios Buza M., Roth E. B., Young J. L. and Drummond M. J., (2023), A bridge to specifications grading in second semester general chemistry, *J. Chem. Educ.*, **100**(6), 2159–2165, DOI: [10.1021/acs.jchemed.2c00731](https://doi.org/10.1021/acs.jchemed.2c00731).
- Nylund K. L., Asparouhov T. and Muthén B. O., (2007), Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study, *Struct. Equ. Model.*, **14**(4), 535–569, DOI: [10.1080/10705510701575396](https://doi.org/10.1080/10705510701575396).
- Nylund-Gibson K. and Choi A. Y., (2018), Ten frequently asked questions about latent class analysis, *Transl. Iss. Psychol. Sci.*, **4**(4), 440–461, DOI: [10.1037/tps0000176](https://doi.org/10.1037/tps0000176).
- Nylund-Gibson K. and Masyn K. E., (2016), Covariates and mixture modeling: results of a simulation study exploring the impact of misspecified effects on class enumeration, *Struct. Equ. Model.*, **23**(6), 782–797, DOI: [10.1080/10705511.2016.1221313](https://doi.org/10.1080/10705511.2016.1221313).
- Nylund-Gibson K., Garber A. C., Carter D. B., Chan M., Arch D. A. N., Simon O., Whaling K., Tartt E. and Lawrie S. I., (2023), Ten frequently asked questions about latent transition analysis, *Psychol. Methods*, **28**(2), 284–300, DOI: [10.1037/met0000486](https://doi.org/10.1037/met0000486).
- Odeleye O. O., Agunbiade O. D., Garber A. and Nylund-Gibson K., (2025), Investigating the evolution of student attitudes toward science in a general chemistry course using latent class and latent transition analysis, *J. Chem. Educ.*, **102**(5), 1745–1754, DOI: [10.1021/acs.jchemed.4c01247](https://doi.org/10.1021/acs.jchemed.4c01247).
- Offerdahl E. G., Hodgson A. and Krupke C., (2016), Lowering the activation barrier, not the academic bar: the role of contract grading in decreasing DFW rates in introductory biochemistry, *FASEB J.*, **30**(S1), 662.619, DOI: [10.1096/fasebj.30.1\\_supplement.662.19](https://doi.org/10.1096/fasebj.30.1_supplement.662.19).
- Partanen L. J., Myrty L. and Asikainen H., (2024), Physical chemistry students' learning profiles and their relation to study-related burnout and perceptions of peer and self-assessment, *Chem. Educ. Res. Pract.*, **25**(2), 474–490, DOI: [10.1039/D3RP00172E](https://doi.org/10.1039/D3RP00172E).
- Patton L. D., (2016), Disrupting postsecondary prose: Toward a Critical Race Theory of higher education, *Urban Educ.*, **51**(3), 315–342, DOI: [10.1177/0042085915602542](https://doi.org/10.1177/0042085915602542).
- Peugh J. and Fan X., (2013), Modeling unobserved heterogeneity using latent profile analysis: a monte carlo simulation, *Struct. Equ. Model.*, **20**(4), 616–639, DOI: [10.1080/10705511.2013.824780](https://doi.org/10.1080/10705511.2013.824780).
- Pintrich P. R. and De Groot E. V., (1990), Motivational and self-regulated learning components of classroom academic performance, *J. Educ. Psychol.*, **82**(1), 33–40, DOI: [10.1037/0022-0663.82.1.33](https://doi.org/10.1037/0022-0663.82.1.33).
- Popova M., (2024), Bridging chemistry education research and practice through research-practice partnerships, *Front. Educ.*, **9**, 1401835, DOI: [10.3389/educ.2024.1401835](https://doi.org/10.3389/educ.2024.1401835).
- Popova M., Shi L., Harshman J., Kraft A. and Stains M., (2020), Untangling a complex relationship: teaching beliefs and instructional practices of assistant chemistry faculty at research-intensive institutions, *Chem. Educ. Res. Pract.*, **21**(2), 513–527, DOI: [10.1039/C9RP00217K](https://doi.org/10.1039/C9RP00217K).
- Popova M., Kraft A., Harshman J. and Stains M., (2021), Changes in teaching beliefs of early-career chemistry faculty: a longitudinal investigation, *Chem. Educ. Res. Pract.*, **22**(2), 431–442, DOI: [10.1039/D0RP00313A](https://doi.org/10.1039/D0RP00313A).
- Pratt J. M. and Raker J. R., (2020), Exploring student affective experiences in inorganic chemistry courses: understanding student anxiety and enjoyment, *Advances in teaching inorganic chemistry volume 1: Classroom innovations and faculty development*, Washington, DC: American Chemical Society, vol. 1370, ch. 10, pp. 117–129, DOI: [10.1021/bk-2020-1370.ch010](https://doi.org/10.1021/bk-2020-1370.ch010).
- Pratt J. M., Stewart J. L., Reisner B. A., Bentley A. K., Lin S., Smith S. R. and Raker J. R., (2023), Measuring student motivation in foundation-level inorganic chemistry courses: a multi-institution study, *Chem. Educ. Res. Pract.*, **24**(1), 143–160, DOI: [10.1039/D2RP00199C](https://doi.org/10.1039/D2RP00199C).
- Pulukuri S. V., Torres D. and Abrams B., (2024), Investigating attitudinal profiles and disparities in attitudes among historically marginalized undergraduate chemistry students, *J. Chem. Educ.*, **101**(9), 3703–3712, DOI: [10.1021/acs.jchemed.4c00640](https://doi.org/10.1021/acs.jchemed.4c00640).
- R Core Team, (2025), *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing.
- Raker J. R. and Holme T. A., (2014), Investigating faculty familiarity with assessment terminology by applying cluster analysis to interpret survey data, *J. Chem. Educ.*, **91**(8), 1145–1151, DOI: [10.1021/ed500075e](https://doi.org/10.1021/ed500075e).
- Raker J. R., Reisner B. A., Smith S. R., Stewart J. L., Crane J. L., Pesterfield L. and Sobel S. G., (2015a), Foundation coursework in undergraduate inorganic chemistry: results from a national survey of inorganic chemistry faculty, *J. Chem. Educ.*, **92**(6), 973–979, DOI: [10.1021/ed500624t](https://doi.org/10.1021/ed500624t).
- Raker J. R., Reisner B. A., Smith S. R., Stewart J. L., Crane J. L., Pesterfield L. and Sobel S. G., (2015b), In-depth coursework in undergraduate inorganic chemistry: results from a national survey of inorganic chemistry faculty, *J. Chem. Educ.*, **92**(6), 980–985, DOI: [10.1021/ed500625f](https://doi.org/10.1021/ed500625f).
- Raker J. R., Gibbons R. E. and Cruz-Ramírez de Arellano D., (2019), Development and evaluation of the organic chemistry-specific achievement emotions questionnaire (AEQ-OCHEM), *J. Res. Sci. Teach.*, **56**(2), 163–183, DOI: [10.1002/tea.21474](https://doi.org/10.1002/tea.21474).
- Renn K. A. and Reason R. D., (2023), *College students in the United States: Characteristics, experiences, and outcomes*, New York, NY: Routledge, DOI: [10.4324/9781003443445](https://doi.org/10.4324/9781003443445).



- Rosenberg J. M., Beymer P. N., Anderson D. J., van Lissa C. J. and Schmidt J. A., (2018), tidyLPA: an R package to easily carry out latent profile analysis (LPA) using open-source or commercial software, *J. Open Source Softw.*, **3**(30), 978, DOI: [10.21105/joss.00978](https://doi.org/10.21105/joss.00978).
- Ryan R. M. and Deci E. L., (2000), Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being, *Am. Psychol.*, **55**(1), 68–78, DOI: [10.1037/0003-066X.55.1.68](https://doi.org/10.1037/0003-066X.55.1.68).
- Schultz M., Lai J., Ferguson J. P. and Delaney S., (2021), Topics amenable to a systems thinking approach: secondary and tertiary perspectives, *J. Chem. Educ.*, **98**(10), 3100–3109, DOI: [10.1021/acs.jchemed.1c00203](https://doi.org/10.1021/acs.jchemed.1c00203).
- Scrucca L., Fraley C., Murphy T. B. and Raftery A. E., (2023), *Model-based clustering, classification, and density estimation using mclust in R*, New York, NY: Chapman and Hall/CRC.
- Sizemore L., Hutchinson B. and Borda E., (2024), Use of machine learning to analyze chemistry card sort tasks, *Chem. Educ. Res. Pract.*, **25**(2), 417–437, DOI: [10.1039/D2RP00029F](https://doi.org/10.1039/D2RP00029F).
- Slominski T., Odeleye O. O., Wainman J. W., Walsh L. L., Nylund-Gibson K. and Ing M., (2024), Calling for equity-focused quantitative methodology in discipline-based education research: an introduction to latent class analysis, *CBE—Life Sci. Educ.*, **23**(4), es11, DOI: [10.1187/cbe.24-01-0023](https://doi.org/10.1187/cbe.24-01-0023).
- Sorich L. and Dweck C. S., (1997), Reliability data for new scales measuring students' beliefs about effort and responses to failure, Unpublished Raw Data, Columbia University.
- Stains M., Harshman J., Barker M. K., Chasteen S. V., Cole R., DeChenne-Peters S. E., Eagan M. K., Esson J. M., Knight J. K., Laski F. A., Levis-Fitzgerald M., Lee C. J., Lo S. M., McDonnell L. M., McKay T. A., Michelotti N., Musgrove A., Palmer M. S., Plank K. M., Rodela T. M., Sanders E. R., Schimpf N. G., Schulte P. M., Smith M. K., Stetzer M., Van Valkenburgh B., Vinson E., Weir L. K., Wendel P. J., Wheeler L. B. and Young A. M., (2018), Anatomy of STEM teaching in North American universities, *Science*, **359**(6383), 1468–1470, DOI: [10.1126/science.aap8892](https://doi.org/10.1126/science.aap8892).
- Vallerand R. J., Pelletier L. G., Blais M. R., Briere N. M., Senecal C. and Vallieres E. F., (1992), The academic motivation scale: a measure of intrinsic, extrinsic, and amotivation in education, *Educ. Psychol. Meas.*, **52**(4), 1003–1017, DOI: [10.1177/0013164492052004025](https://doi.org/10.1177/0013164492052004025).
- van Driel J. H., Bulte A. M. W. and Verloop N., (2005), The conceptions of chemistry teachers about teaching and learning in the context of a curriculum innovation, *Int. J. Sci. Educ.*, **27**(3), 303–322, DOI: [10.1080/09500690412331314487](https://doi.org/10.1080/09500690412331314487).
- Velasco J. B., Knedeisen A., Xue D., Vickrey T. L., Abebe M. and Stains M., (2016), Characterizing instructional practices in the laboratory: the laboratory observation protocol for undergraduate STEM, *J. Chem. Educ.*, **93**(7), 1191–1203, DOI: [10.1021/acs.jchemed.6b00062](https://doi.org/10.1021/acs.jchemed.6b00062).
- Vermunt J. K., (2017), Latent class modeling with covariates: two improved three-step approaches, *Political Anal.*, **18**(4), 450–469, DOI: [10.1093/pan/mpq025](https://doi.org/10.1093/pan/mpq025).
- Villalta-Cerdas A. and Sandi-Urena S., (2014), Self-explaining effect in general chemistry instruction: eliciting overt categorical behaviours by design, *Chem. Educ. Res. Pract.*, **15**(4), 530–540, DOI: [10.1039/C3RP00172E](https://doi.org/10.1039/C3RP00172E).
- Wang Y. and Lewis S. E., (2022), Towards a theoretically sound measure of chemistry students' motivation; investigating rank-sort survey methodology to reduce response style bias, *Chem. Educ. Res. Pract.*, **23**(1), 240–256, DOI: [10.1039/D1RP00206F](https://doi.org/10.1039/D1RP00206F).
- Wang M., Sundstrom M., Nylund-Gibson K. and Ing M., (2025a), Advancing clustering methods in physics education research: a case for mixture models, *Phys. Rev. Phys. Educ. Res.*, **21**(2), 020126, DOI: [10.1103/1fn4-nqvj](https://doi.org/10.1103/1fn4-nqvj).
- Wang Y., Machost H., Yik B. J. and Stains M., (2025b), Why chemistry instructors are shifting to specifications grading: perceived benefits and challenges, *Chem. Educ. Res. Pract.*, **26**(4), 846–866, DOI: [10.1039/D5RP00035A](https://doi.org/10.1039/D5RP00035A).
- Wilkes C. L., Gamble M. M. and Rocabado G. A., (2024), Is general chemistry too costly? How different groups of students perceive the task effort and emotional costs of taking a chemistry course and the relationship to achievement and retention, *Chem. Educ. Res. Pract.*, **25**(4), 1090–1104, DOI: [10.1039/D4RP00034J](https://doi.org/10.1039/D4RP00034J).
- Xu X. and Lewis J. E., (2011), Refinement of a chemistry attitude measure for college students, *J. Chem. Educ.*, **88**(5), 561–568, DOI: [10.1021/ed900071q](https://doi.org/10.1021/ed900071q).
- Yik B. J., Machost H., Streifer A. C., Palmer M. S., Morkowchuk L. and Stains M., (2024), Students' perceptions of specifications grading: development and evaluation of the Perceptions of Grading Schemes (PGS) instrument, *J. Chem. Educ.*, **101**(9), 3723–3738, DOI: [10.1021/acs.jchemed.4c00698](https://doi.org/10.1021/acs.jchemed.4c00698).
- Yik B. J., Morkowchuk L., Wheeler L. B., Roksa J., Machost H. and Stains M., (2025), Balancing equity in general chemistry laboratory courses: the complex impact of specifications grading on student success and opportunity gaps, *JACS Au*, **5**(6), 2593–2605, DOI: [10.1021/jacsau.5c00210](https://doi.org/10.1021/jacsau.5c00210).
- Zhang Y. and Yik B. J., (2026), K-means clustering & latent profile analysis comparison, OSF, DOI: [10.17605/OSF.IO/TAKR4](https://doi.org/10.17605/OSF.IO/TAKR4).

