Chemistry Education Research and Practice

Design, development, and evaluation of the Organic chemistry Representational Competence Assessment (ORCA)

Journal:	*Chemistry Education Research and Practice*
Manuscript ID	RP-ART-07-2023-000188.R4
Article Type:	Paper

ARTICLE

Design, development, and evaluation of the Organic chemistry Representational Competence Assessment (ORCA)

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Lyniesha Ward,[a] Fridah Rotich,[a] Jeffrey R. Raker,[b] Regis Komperda,[c] Sachin Nedungadi,[d] and Maia Popova*[a]

This paper describes the design and evaluation of the Organic chemistry Representational Competence Assessment (ORCA). Grounded in Kozma and Russell's representational competence framework, the ORCA measures the learner's ability to *interpret*, *translate,* and *use* six commonly used representations of molecular structure (condensed structures, Lewis structures, skeletal structures, wedge-dash diagrams, Newman projections, and chair conformations). Semi-structured interviews with 38 first-semester organic chemistry learners informed the development of the ORCA items. The ORCA was developed and refined through three pilot administrations involving a total of 3,477 first-semester organic chemistry students from multiple institutions. The final version of the ORCA was completed by 1,494 students across five institutions. Various analyses provided evidence for the validity and reliability of the data generated by the assessment. Both one-factor and three-factor correlated structures were explored via confirmatory factor analysis. The one-factor model better captured the underlying structure of the data, which suggests that representational competence is better evaluated as a unified construct rather than as distinct, separate skills. The ORCA data reveal that the representational competence skills are interconnected and should consistently be reinforced throughout the organic chemistry course.

Introduction and rationale

Learning and communicating with representations (*e.g.*, chemical structures, graphs, or reaction equations) are essential goals of chemistry instruction (Ainsworth, 2006; Gilbert, 2005; Kozma & Russell, 2005). Correspondingly, representations are integral to chemistry assessment. A historical analysis of 18 organic chemistry American Chemical Society (ACS) exams revealed that more than 90% of exam items include at least one representation (Raker & Holme, 2013). The ACS Exams Institute includes visualizations as one of the ten anchoring concepts for undergraduate chemistry (Murphy et al., 2012). Additionally, the National Academies of Science emphasize the importance of reasoning with representations to engage in scientific and engineering practices (*e.g.*, developing and using models, analysing and interpreting data, and obtaining, evaluating, and communicating information) (National Research Council, 2012a). It is, therefore, unsurprising that science educators and discipline-based education researchers have devoted considerable attention to learners' understanding of scientific representations (National Research Council, 2012b).

Mastering the "language" of chemical representations is not an easy task. Novices must learn *about* representations (*e.g.*, how to interpret representations) and learn *with* representations (*e.g.*, use representations to make sense of chemical phenomena) (Talanquer, 2022). Research shows that learners rely on heuristics when making sense of representations (Talanquer, 2014), primarily focus on surface features of representations (Cooper et al., 2010; Popova & Bretz, 2018c), and struggle to connect the external features of representations to the conceptual information embedded in them (Popova & Bretz, 2018a, 2018b; Rotich et al., 2024; Ward et al., 2022). Despite the important role that representational competence plays in learner success in chemistry (Ainsworth, 2006; Kozma & Russell, 2005; Schönborn & Anderson, 2008; Seufert, 2003; Stieff et al., 2016), and the considerable number of investigations into learners' ability to reason about symbolic and submicroscopic representations (Bodner & Domin, 2000; Connor et al., 2019; Cooper et al., 2010; Farheen and Lewis, 2021; Farheen et al., 2024, Lansangan et al., 2018; Miller & Kim, 2017; Padalkar & Hegarty, 2012, 2015; Popova & Bretz, 2018a, 2018b, 2018c; Stieff et al., 2016; Stull et al., 2016; Stull & Hegarty, 2016; Ward et al., 2022; Wright & Oliver-Hoyo, 2020), a limited number of assessment instruments exist that measure learner competence with representations. Within this manuscript, we describe the development of an assessment to measure aspects of learner representational competence in the context of multiple commonly used representations of molecular structure.

Conceptual framework – Representational competence

[a] University of North Carolina at Greensboro, Greensboro, North Carolina, USA. *m_popova@uncg.edu
[b] University of South Florida, Tampa, Florida, USA
[c] San Diego State University, San Diego, California, USA
[d] University of Nebraska Omaha, Omaha, Nebraska, USA

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

Representational competence (RC) is a "set of skills and practices that allow a person to reflectively use a variety of representations or visualizations, singly and together, to think about, communicate, and act on chemical phenomena in terms of underlying, aperceptual physical entities and processes" (Kozma and Russell, 2005, p. 131). Through an examination of expert and novice investigative chemistry practices (Kozma et al., 2000; Kozma & Russell, 1997), Kozma and Russell (2005) proposed a set of skills that comprise RC:

(a) use representations to describe observable chemical phenomena in terms of underlying molecular entities and processes (*use*),
(b) generate or select a representation and explain why it is appropriate for a particular purpose (*generate and select)*,
(c) use words to interpret features of a particular representation (*interpret)*,
(d) make connections across different related representations by mapping features of one representation onto those of another (*translate)*,
(e) use representations in social situations to support claims, draw inferences, and make predictions about chemical phenomena (*use)*,
(f) describe the limitations and affordances of different representations (*identify affordances and limitations*), and
(g) take the epistemological position that representations correspond to but are distinct from the phenomena that are observed (*take epistemological position)*.

Most skills outlined above can be developed at a lower foundational level or a higher meta-level that allows for the reflective and purposeful use of representations (diSessa, 2004; diSessa, Andrea & Sherin, 2000; Gilbert, 2005; Kozma & Russell, 2005; Ward et al., 2022). For the remainder of the article, we will use the succinct labels in parentheses to reference each skill.

RC is grounded in the cognitive theory of multimedia learning and situative theory. Together, these two theories highlight the interplay between cognitive processing and the social context of learning for understanding how students master chemical representations. The cognitive theory of multimedia learning assumes that learning occurs as a result of processing and synthesizing information across the audio and visual modes of instructional information (Kozma & Russell, 2005; Mayer, 2002). In particular, the ability to effectively process visual information is critical when it comes to learning chemical representations, as these often involve intricate visual details that convey essential information about sub-microscopic particles and processes. Situative theory suggests that the physical and social characteristics of a setting shape the processes of understanding and learning within that setting (Kozma & Russell, 2005; Lave, 1991; Lave & Wenger, 1991). As individuals become integrated into a community, they progressively develop in using its representational systems to construct new knowledge and communicate information. This integration and participation in communal practices enables learners to understand and manipulate the representations used by that community, facilitating deeper learning and expertise in the subject matter.

Several assumptions underlie the development of RC. First, a developmental trajectory is assumed, implying that learners' ability to use chemistry representations increases as they progress in their learning. With continued learning, representations become a valuable tool for constructing and communicating understanding, resulting in the sophistication of RC over time (Kozma & Russell, 2005). Second, levels of RC are not assumed to be consistent for different types of representations. For example, a learner may be proficient with most of the RC skills for a particular representation (such as Lewis structures) and only be able to interpret the features of another (such as reaction coordinate diagrams) (Halverson & Friedrichsen, 2013; Kohl & Finkelstein, 2006; Stieff et al., 2011). Finally, RC and conceptual understanding are separate but related components of learner knowledge (Hinze et al., 2013; Maroo & Johnson, 2018). Proficiency in understanding chemical concepts is achieved in tandem with proficiency in utilizing the visualizations that represent and explain those concepts (diSessa, 2004; Kozma & Russell, 2005).

Studies investigating learner RC in chemistry have primarily relied on the use of interviews. Interviews provide nuanced and rich data on how learners engage and reason with representations that result in transferable conclusions. Many of these studies suggested a relationship between RC and conceptual understanding (Hiong & Daniel, 2015; Höst et al., 2012; Lansangan et al., 2007, 2018; Pande et al., 2015; Pande & Chandrasekharan, 2022). However, interview data do not allow for drawing robust, generalizable conclusions for various groups of learners, for example, those experiencing different learning environments (Herrington & Daubenmire, 2014). Studies that use assessments that can be easily administered and scored on a large scale have the potential to improve our understanding of RC as a construct and how to effectively develop RC through instruction. This need was highlighted by Kozma and Russell, who noted that "new assessments must be designed and used that measure investigation practices and related skills, such as visualization skills or representational competence" (Kozma & Russell, 2005, p. 142).

RC (Connor et al., 2021; Sim & Daniel, 2014), or frameworks related to RC (e.g., Visualization Competence of Matter (Chang & Tzeng, 2017) and Visual Model Comprehension (Dickmann et al., 2019)) have been the foundation of some assessments developed to elicit secondary or postsecondary chemistry learners' ability to make sense of representations. These assessments have corroborated findings from interviews stating that representational abilities are greater for learners who received more chemistry instruction (Chang & Tzeng, 2017; Vlacholia et al., 2017) or who have a higher conceptual understanding (Dickmann et al., 2019; Sim & Daniel, 2014). In addition, a few studies found that competence varies by skill or topic (Chang, 2018; Chang & Tzeng, 2017). However, these assessments are either not used to investigate learner RC explicitly (Chang, 2018; Chang & Tzeng, 2017; Dickmann et al., 2019; Vlacholia et al., 2017; Wang et al., 2017), are focused on a single aspect of RC (Connor et al., 2021), or a limited number of representations (Connor et al., 2021; Sim & Daniel, 2014). Some of these assessments have useful but limited evidence for the validity of resulting data, primarily in the form of content validity using expert panels (Chang, 2018; Chang & Tzeng, 2017). This contrasts with more robust evaluations of validity and reliability, including content validity as well as a broader array of psychometric evidence. To advance our understanding of learner RC, quality assessments that demonstrate evidence of valid and reliable data are needed that comprehensively capture multiple RC skills in the context of multiple representations (Kozma & Russell, 2005).

Purpose of this study

This work aims to develop and evaluate the Organic chemistry Representational Competence Assessment (ORCA) to characterize learners' RC in the context of six commonly used representations of molecular structure in undergraduate organic chemistry instruction: condensed structures, chair conformations, wedge-dash diagrams, Lewis structures, skeletal structures, and Newman projections. Chemical representations have various dimensions (*i.e.*, iconicity, granularity, dimensionality, and quantitativeness) that impact how learners reason about and with representations (Talanquer, 2022). This study, in particular, centers on symbolic representations with similar dimensions (Johnstone, 1993).

The ORCA is designed as a multiple-choice assessment, permitting testing with a large number of participants and allowing for quick grading. The assessment captures three RC skills: the ability to (a) use words to interpret features of a particular representation – the *interpret* skill; (b) make connections across different related representations by mapping features of one representation onto those of another – the *translate* skill; and (c) use representations to draw inferences – the *use* skill. These skills are among the most commonly taught by organic chemistry educators (Jones et al., 2022; Linenberger & Holme, 2015; Popova & Jones, 2021) and in organic chemistry textbooks (Gurung et al., 2022). The ORCA does not assess the learner's ability to *generate* representations, another skill typically taught in the target course, as it is difficult to evaluate this skill properly using multiple-choice questions. Other RC skills were not considered for the ORCA as current instruction offers little to no support for those skills (Gurung et al., 2022; Jones et al., 2022; Linenberger & Holme, 2015; Popova & Jones, 2021). Two research questions guide this study:

1. What evidence exists for the validity and reliability of the data generated from the Organic chemistry Representational Competence Assessment (ORCA)?
2. What do the data collected with ORCA reveal about learner representational competence?

Methods: Data collection, analysis, and evidence of validity and reliability of the data

Criteria for developing and evaluating the ORCA were adopted from the Standards for Educational and Psychological Testing (American Educational Research Association, 2014) and guided by other primers for assessment development in chemistry education (Arjoon et al., 2013; Komperda et al., 2018; Towns, 2008, 2014). The development and evaluation processes are organized by the various stages and shown in **Table 1.** Institutional Review Board guidelines were met at every institution where data were collected and in all stages of the assessment development and evaluation. All participants were informed that their ORCA performance would not negatively impact their course grades. Additionally, the participants could consent or decline consent for their responses to be included in this research study.

We iteratively developed the ORCA by using a sequential, mixed-methods exploratory design (Towns, 2008) and a bottom-up approach that originated from analyses of learner reasoning about representations of molecular structure collected using semi-structured interviews (Bowen, 1994). At the time of their participation, all participants were in the first semester of a year-long organic chemistry course, having completed the two-semester general chemistry sequence. We wrote assessment items from the interview data collected at one institution (Stage I), and we refined these items through three pilot administrations with participants at six institutions (Stage II) using the Qualtrics platform. The final ORCA version was administered to learners at five institutions (Stage III) via Qualtrics. Only learners who completed the entire assessment received extra credit for their participation, which is why all participants completed ORCA in its entirety, and we did not have any missing data.

Stage I – Assessment development

We conducted semi-structured interviews; data from a portion of these interviews were used to design the ORCA items. We recruited participants (N = 38) from a public university in the southeastern US during the Fall of 2019 and Spring of 2020. The participants were enrolled in five different Organic Chemistry I course sections taught by four instructors using the same curriculum and textbook (Bruice, 2016). We purposefully sampled learners, using a stratified method (Patton, 2002), to ensure a range of grades (from "A" to "C") in General Chemistry II. Participants were compensated with a $20 gift card. Pseudonyms were assigned to all participants to protect their identities. Data were captured using an audio recorder, a video camera, and a Livescribe smartpen (http://www.livescribe.com).

The interviews were designed to elicit learners' RC skills. To capture how learners *interpret* a representation, participants were asked to (a) describe what the representation communicates, (b) decode each diagrammatic feature of the representation (*e.g.*, dashes, wedges, lines, symbols, etc.), (c) explain the purpose of the representation, and (d) identify the atomic composition for the given representation. To elicit how learners *use* a representation, participants were asked to examine a given representation and make inferences about (a) chemical and physical properties, (b) bonding, and (c) energy and stability. To capture how learners *translate* between various representations of molecular structure, the participants were given a specific representation and asked to choose the corresponding structure(s) from four options. They were informed that each task may have more than one corresponding structure. After selecting the corresponding structure(s), the participants were asked to explain why they believed the structure(s) corresponded and why the other options did not.

Table 1. ORCA development and evaluation.

Method	Outcome	Application
Stage I – Assessment development		
Semi-structured think-aloud interviews	Elicit learner ability to reason with representations	Generate items and distractors from participant responses to establish ***validity evidence based on test content***
Expert evaluation	Obtain feedback from experts about the appropriateness of the items for the measurement of the target construct	Establish ***validity evidence based on test content*** and revise items accordingly
Stage II – Pilot administrations		
Item analysis	Understand item difficulty, item discrimination, and how well distractors function	Revise items accordingly to ensure ***item quality***
Factor analysis	Investigate the number of factors, the variables that load onto each factor, and the level of correlation among factors	Establish ***internal structure validity*** and revise instrument accordingly
Cognitive interviews	Obtain insight into respondents' thought processes and feedback about items	Establish ***response process validity*** and revise items accordingly
Expert evaluation	Obtain additional feedback from experts about the appropriateness of the items for the measurement of the target construct	Establish ***validity evidence based on test content*** and revise items accordingly
Stage III – Final administration		
Item analysis	Understand item difficulty, item discrimination, and how well distractors function	Ensure ***item quality***
Factor analysis	Confirm the factor structure of the revised ORCA instrument	Establish ***internal structure validity***
Reliability coefficient	Understand correlations between items	Establish ***internal consistency/reliability***

The data from the interviews were transcribed verbatim, inductively coded using ATLAS.ti software (version 9, https://atlasti.com/), and analyzed using constant comparative analysis (Glaser, 1965). The first and the second author coded the data and, depending on a code, obtained 84-100% agreement (Saldaña, 2013). Furthermore, they discussed each case of disagreement in their coding until a 100% negotiated agreement was reached. In addition, the first, second, and sixth authors met weekly to discuss the codes and analyze the data for patterns. Some of the qualitative results from these analyses have been published elsewhere (Rotich et al., 2024; Ward et al., 2022).

We then used these data to generate over 100 assessment items to capture how learners *interpret*, *translate*, and *use* six representations of molecular structure. **Figure 1** depicts how we used participant interview data (**Figure 1A**) to write an item that asks participants to *use* Newman projections to make inferences about stability (**Figure 1B**). Response choices, including distractors, were developed using participants' ideas from the interview and reflected the most common patterns in participant thinking. We reviewed all the items and eliminated items if: (a) there were too many items about the same representation or skill, (b) items were complex multiple-choice questions (Albanese, 1993; Towns, 2014), or (c) items were not representative of the skills learners typically learn in the course (*i.e.*, *translating* between a chair conformation and a condensed structure). This process resulted in condensing the assessment to 64 items.

An expert panel of six organic chemistry instructors and chemistry assessment developers evaluated all 64 items for content validity. We used the feedback from the expert panel members (**Figure 1C**) to eliminate some items and refine other items (**Figure 1D**). This process resulted in 48 items for which the expert panel established that the items adequately covered the relevant content to measure the target constructs (*interpret, translate* between, and *use* representations of molecular structure) and were appropriate for the target population (organic chemistry learners). In addition, they provided feedback about terminology and the scoring of items.

Stage II – Pilot administrations

Following expert evaluation, we pilot-tested the 48 items. During each pilot test, participants completed the assessment at least three weeks after covering the six target representations of molecular structure in their Organic Chemistry I course. In consensus with the feedback received from the expert panel, and to minimize item priming, we administered the assessment items in a specific order, such that *translate* items appeared first, then *use* items, and lastly, *interpret* items. Within each skill, the items were randomized for each participant.

Pilot I: In Fall 2021, the 48 items were administered as a Pilot I to learners (N = 1,120) taught by five instructors at two (medium and large) public universities in the southeastern US. To ensure a timely assessment, we split the items into two 24-item forms that could be completed in less than 30 minutes. Each form had comparable items that featured the same representations and targeted the same skills. For example, Form 1 and Form 2 each contained one item where participants were to *use* a Newman projection to make inferences about stability. Participants randomly received one of the two

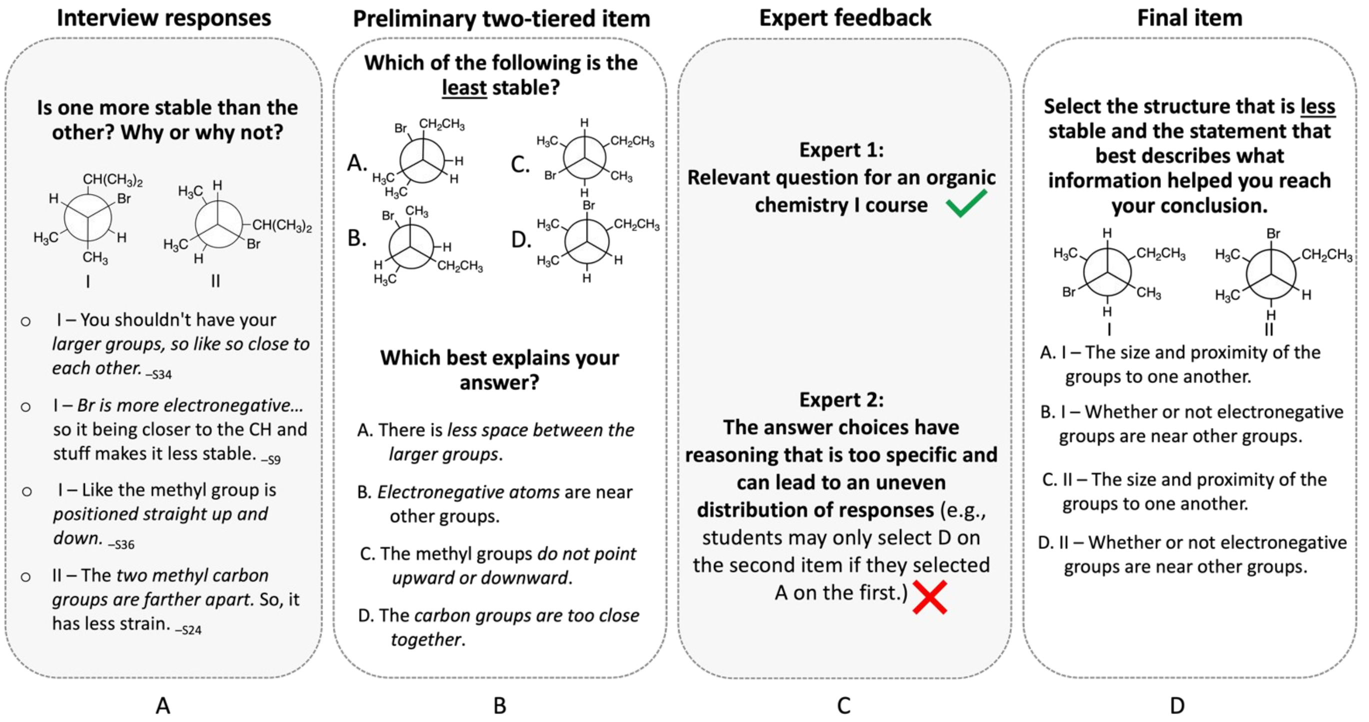


Figure 1. An example preliminary item (B) written based on participant interview data (A), as well as the refined item (D) based on expert feedback (C) to establish validity evidence based on test content.

comparable forms. We used classical test theory considerations (*i.e.*, item difficulty, discrimination, and response choice distributions) (Bandalos, 2018; Towns, 2014) to select the best performing items across the two forms. This resulted in a 24-item assessment with seven *interpret* skill items, eleven *translate* skill items, and six *use* skill items. The six *use* items asked participants to infer stability from Newman projections, chair conformations, and Lewis structures (*i.e.*, two items for each representation).

Pilot II: We administered the 24 items as a Pilot II in Spring 2022 to participants (N = 567) at the same two institutions as in Pilot I. This administration was used to provide preliminary evidence of the internal structure of the assessment through confirmatory factor analysis (CFA). Classical test theory considerations revealed the assessment demonstrated a range of item difficulties and discriminations for most items and a few items under the *translate* skill had poor answer choice distribution. We used data from cognitive interviews to improve those items. In addition, we wrote six new *use* items (generated using the interviews from Stage I) to target how learners infer physical properties from condensed structures, skeletal structures, and wedge-dash diagrams. We incorporated these new items into the response process interviews as well. We conducted cognitive interviews with learners (N = 5) two weeks after the Pilot II administration. We used stratified purposeful sampling to select high, middle, and low-scoring ORCA participants to obtain evidence for response process validity (Deng et al., 2021; Patton, 2002). For each item, the participants responded to the following questions:

- What does this question ask you to do?
- What information do you believe is relevant for responding to this question?
- Generate a representation that corresponds to this structure (*translate* items only).
- How did you decide to pick <response option selected>?
- Did you experience any challenges when providing your response to this question?

During the interview, participants first addressed poor-performing items from the assessment and then the six newly written items. We used responses from the interviews to refine the item stems, modify distractors, and add or delete items. For some items, participants pointed out key features of the representations that cued them to eliminate a distractor or made the correct answer obvious. Based on participants' feedback, we made adjustments to these items, deleted two items where participants described confusing terminology, and added four of the six newly generated *use* items to the assessment.

Pilot III: Finally, we tested the finalized 26-item ORCA again as a Pilot III in the middle of the Fall 2022 semester with participants (N = 1,790) across six medium and large public universities in the southeastern, midwestern, and western US. The instrument contained seven *interpret* items, eleven *translate* items, and eight *use* items. Once again, we examined the quality of the items for distractor selection, item difficulty, and item discrimination, as well as conducted CFA to investigate the internal structure of the assessment. Only one item performed poorly on Pilot III. The terminology for this *use* item had been adjusted after the response process interviews in Pilot II and this led to poor discrimination and low difficulty. Recognizing the impact of this change, we reverted the terminology of this item to its original format from Pilot II for the final administration. Additionally, the expert panel evaluated the refined assessment instrument to confirm the content validity. Persons interested in obtaining a copy of the finalized ORCA should see the ESI.

Stage III – Final administration

The final ORCA was administered at the end of Fall 2022 to Organic Chemistry I learners from five institutions: one medium public university in the southeastern US and four large public universities in the southeastern, midwestern, and western US. Participants completed the assessment with a median time of 22 minutes, and 1,508 participants consented for their ORCA responses to be used in this study. Many participants in Pilot III were also part of the final administration. However, the timeframe between the two administrations was over six weeks to minimize the likelihood of participants recalling specific items from Pilot III. Additionally, to further reduce the possibility of recall effects, we implemented a randomization strategy for both the items and their response choices during the final administration. This means that even if students participated in both phases, the presentation of the items and response options in the final administration differed from their arrangement in Pilot III.

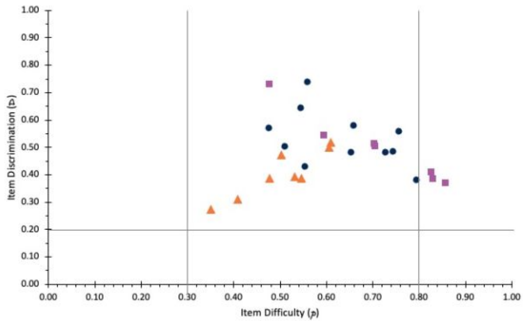
Next, we identified any potential data outliers. Based on the recommendations of Aguinis and colleagues (2013), we used single (*i.e.*, box plots and percentage analysis) and multiple construct techniques (*i.e.*, scatterplots, Mahalanobis distance, and studentized deleted residuals). Fourteen outliers (*i.e.*, complete response sets for a given participant) were removed through these analyses. Outlier data had *interpret* item scores of zero and often had extremely high (*e.g.*, 340 minutes) or low (*e.g.*, 94 seconds) completion times. This process resulted in a final sample of 1,494 responses for analysis.

We conducted CFA using MPlus 8 (Muthén & Muthén, 2017) to investigate the internal structure of the assessment instrument. With the RC framework (specifically, the *interpret, translate,* and *use* skills) as a theoretical basis, we tested two hypothesized models to evaluate and confirm the number and nature of latent factors. The Means and Variance Adjusted Weighted Least Squares (WLSMV) method was best suited for the analysis (Finney & DiStefano, 2013) as all measured variables were categorical (binary). A good model fit for these categorical data is more stringent than the criteria suggested by Hu and Bentler (Beauducel & Herzberg, 2006; Hu & Bentler, 1999; McNeish & Wolf, 2023; Xia & Yang, 2019). Though the model fit cutoffs are not absolute and the conditions for their use are still undergoing research (McNeish et al., 2018; McNeish & Wolf, 2023), we use criteria adopted by Komperda and colleagues for good fit (*i.e.*, Comparative Fit Index (CFI) > .95 and the Root Mean Square Error of Approximation (RMSEA) < .05). The proposed models were congeneric, which is why McDonald's omega (ω) was calculated as a reliability indicator to examine the internal consistency of the items within each latent factor (Hancock and An, 2018; Komperda, Pentecost, and Barbera, 2018).

Results and discussion

Gathering evidence for the internal structure validity and internal consistency of the ORCA data

Item difficulty and discrimination from the final ORCA administration with 1,494 participants are reported in **Figure 2.**



All items sufficiently discriminated (ρ > 0.2) between the top 27% and bottom 27% of performers (Bandalos, 2018). There was a range of difficulties for the items, with most falling in the 0.3 – 0.8 range.

Figure 2. Item difficulty and discrimination for the final ORCA administration. *Interpret* skill items are shown as purple squares, *translate* skill items as blue circles, and *use* skill items as orange triangles.

Confirmatory Factor Analysis. Based on Kozma and Russell's RC framework, we proposed two models to estimate the internal structure validity of the data. Model A is a three-factor model where each latent factor represents an RC skill (*i.e.*, the ability to *interpret*, *translate*, and *use* representations). In this model, each item was set to load onto its corresponding factor with correlations between the three factors (**Figure 3A**). Model B is a one-factor model in which all

ARTICLE

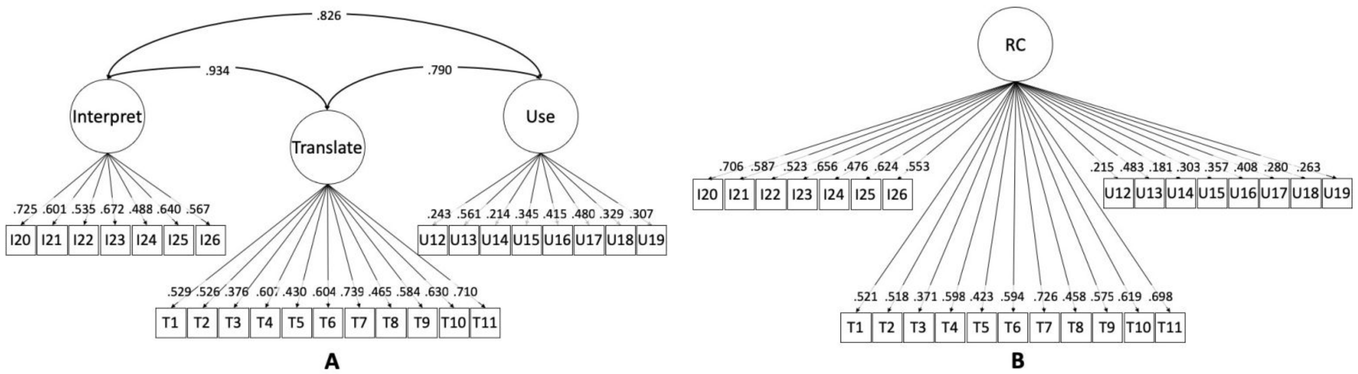


Figure 3. Standardized parameter estimates for the three-factor (A) and one-factor (B) models. Circles indicate the latent variables. Squares indicate the observed variable item scores. Double-headed arrows between circles represent correlations between factors and the arrows between circles and squares represent factor loadings. Items in model A were set to load on their assigned factors only. All factor loadings are significantly different from zero (p < .001).

Table 2. Fit statistics and reliability for the two proposed models.

	Fit Statistics					McDonald's Omega (ω)		
Model	χ^2	df	p	RMSEA ($\leq$ .05)	CFI ($\geq$ .95)	Interpret	Translate	Use
A – 3 factor	694	296	<.0001	.030	.955	.807	.843	.554
B – 1 factor	741	299	<.0001	.031	.950		.897	

items load onto a single latent factor representing RC (**Figure 3B**). Model fit statistics and the McDonald's omega coefficients are reported in **Table 2**.

The data meet the fit criteria for Model A. However, internal consistency for the factors, as measured through McDonald's omega reliability coefficient, varied by skill. The *translate* and *interpret* factors have a higher reliability, indicating that common constructs (*i.e.*, the ability to *translate* between representations and the ability to *interpret* features of an individual representation, respectively) explain a larger amount of the observed variance to the total variance (Komperda, Pentecost, and Barbera, 2018). The *interpret* skill and *translate* skill items require learners to reason by attending to the diagrammatic features within single or multiple representations (*e.g.*, interpreting dashes and wedges in wedge-dash diagrams, mapping dashes and wedges onto the appropriate axial and equatorial substituents in chair conformations). For this reason, the *interpret* skill and *translate* skill factors are highly correlated (**Figure 3A**), as students reason about the features of the representations (i.e., composition, connectivity, or spatial information, if applicable) to *interpret* and *translate* between representations.

In contrast, the *use* skill factor has a lower McDonald's omega, so a common construct (*i.e.*, the ability to *use* representations to make inferences) explains less of the observed variance. This is not surprising because the *use* skill items require participants to not only reason by attending to the features of representations but also by extracting the relevant domain-specific conceptual knowledge embedded in the representation. Moreover, this conceptual knowledge varies by representation (*e.g.*, evaluating the arrangement of substituents in Newman projections to make an inference about *stability*; evaluating the functional groups in skeletal structures to make an inference about *physical properties*). In this case, not only do learners need to be able to make sense of the features of a representation, but they also need to have a conceptual understanding of stability or physical properties, and know which relevant features of the representation to attend to make inferences about these concepts. Therefore, a lower reliability for the *use* skill factor is hypothetically expected as it may not be a homogenous construct (Taber, 2018); however, this lower reliability also limits the inferences that can be made about the *use* skill from the items within this assessment. Moreover, the factor loadings (**Figure 3A**) for the *interpret* and *translate* factors are higher than those in the *use* factor, further indicating that the *interpret* and *translate* skill factors explain more variance within the items (Bandalos, 2018).

Komperda and colleagues (2018) suggest that when "an instrument is known to be composed of multiple scales where scores will be reported separately, each scale should be evaluated to determine if it fits a single-factor model, and a reliability value should be provided for each scale of data." Data for each factor of Model A were evaluated with individual latent factor models; these CFA results provide support that *interpret* and *translate* scales fit the data and have factor loadings > 0.2 (see **Figure S1** and **Table S1** in the Appendix). However, while the RMSEA is appropriate, the CFI for the *use* skill is below the more stringent cut-off for categorical data. Given that the *use* scale has a weaker model fit, these analyses provide evidence that only the *interpret* and *translate* scales can function independently. In summary, although Model A has a solid theoretical basis, aligned with the stance that RC is a "set of skills and practices that allow a person to reflectively use a variety of representations" (emphasis added; Kozma and Russell, 2005, p. 131), and the data meet the thresholds for the fit statistics, the *use* scale has items that have lower factor loadings, weaker internal consistency, and, unlike the *interpret* and *translate* scales, cannot be used independently.

In comparison, Model B is a single-factor model where every item loads onto a single latent factor—RC. The data fit Model B (**Figure 3B**), suggesting that this model effectively captures the underlying structure of the data. Model B accounts for the high latent correlation between the *interpret* and *translate* skill factors (evidenced as 0.934 in **Figure 3A**). While it is possible that the ORCA may be limited in distinguishing these factors, it is more likely that these skills are very related and should be considered one factor, as reflected in Model B. Additionally, the internal consistency (as measured by McDonald's omega) is higher for the one-factor model B than the three-factor model A. While the increased internal consistency may be attributed to the larger number of items within a single factor (Malkewitz et al., 2023), the empirical data better align with the single-factor structure. This suggests that in the context of representations of molecular structure, these RC skills are better evaluated as a unified construct rather than as distinct, separate skills. Below, we delve deeper into the implications of these findings and share insights into measuring student RC.

Insights gained about learner representational competence

The section below outlines two main insights from examining the data collected with the ORCA about student RC with six key representations of molecular structure in undergraduate organic chemistry: condensed structures, chair conformations, wedge-dash diagrams, Lewis structures, skeletal structures, and Newman projections. These representations are essential to student success in organic chemistry, and our findings provide insights into learners' proficiency with these symbolic representations.

Representational competence skills are interconnected. The high correlation between each skill in Model A led to the selection of Model B as the best data-fit model (**Figure 3**). These correlations indicate that the items intended to measure separate skills may actually be assessing the same underlying construct. In practice, it is difficult to *interpret*, *translate*, or *use* representations of molecular structure independently, as these skills inherently rely on and reinforce one another. This is particularly the case with the *interpret* skill, which is requisite for the co-development of the rest of the skills. This interconnectedness suggests that RC might be better understood as an interconnected set of skills rather than distinct, separate abilities. Furthermore, many tasks and assessments in chemistry require learners to simultaneously apply multiple RC skills, reinforcing the idea that these skills are not isolated but are used in concert. This perspective highlights that the effective use of chemical representations often depends on the seamless integration of various interrelated skills.

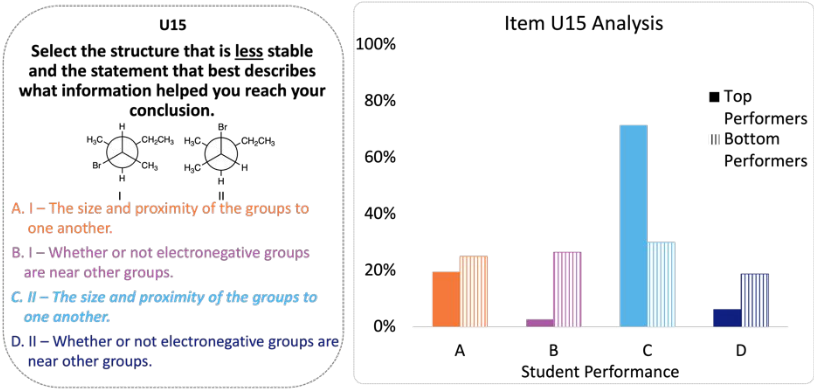
Two examples from our item-level analysis of participant performance further illustrate the interconnectedness of RC skills. Specifically, we observed instances of students having the necessary conceptual understanding to *use* a representation to make an inference but not being able to connect this knowledge to the appropriate features of representations, which relates to the *interpret* skill.

Item U15, shown in **Figure 4A** (the same item as in **Figure 1D**), required participants to make inferences about the stability of Newman projections and justify their reasoning by relying on the concepts embedded in the diagrammatic features of the representation. With the answer choices in this item, participants could have relied on the concepts of substituent size or electronegativity. **Figure 4B** demonstrates that the likelihood of selecting the correct answer C is lower for bottom ORCA performers (bottom 27%) and higher for top performers (top 27%). Almost a quarter of participants (23%, **Figure 4C**) selected response choice A, in which they chose the wrong structure while reasoning about the appropriate concept of substituent size. These participants knew the productive conceptual information to make inferences about stability but could not connect that knowledge to the proper diagrammatic features of the Newman projection. Finally, almost a quarter of the participants relied on the wrong conceptual information (electronegativity) and selected responses D (13%) or B (11%). This quarter of participants comprise less than 9% of the top performers but over 45% of the bottom performers. For this item, the bottom performers struggled to identify the productive concept of interest.

Item U14 (**Figure 5A**) demonstrated a similar pattern. The correct answer, B, was chosen by 41% of all participants. A slight majority of bottom performers (55%) relied on the productive concept reflected in choices B and D. These participants were able to identify the appropriate feature (charge) and connect it to the appropriate concept (octet). However, half of them could not connect that knowledge to the appropriate diagrammatic features to make accurate inferences about stability. Thus, they incorrectly selected the structure with an atom with an incomplete octet (D). Almost a third (30%) of all participants selected response choice D. Finally, nearly a third of the participants chose responses A (20%) or C (9%). These participants relied on features like the presence or absence of double bonds that were not as relevant to the concept of interest (stability) in this prompt.

These examples highlight the complex interplay of RC skills required to answer such questions. Successfully *using* representations to make inferences about a concept demands a combination of knowledge and ability to (1) *interpret* diagrammatic features of a representation, (2) conceptually understand the concept of interest, (3) distinguish the relevant from irrelevant features of the representation in relation to the concept of interest, and (4) connect the relevant features of the representation to one's conceptual understanding to make an inference about the concept of interest. Each of these steps is necessary when making inferences about representations.

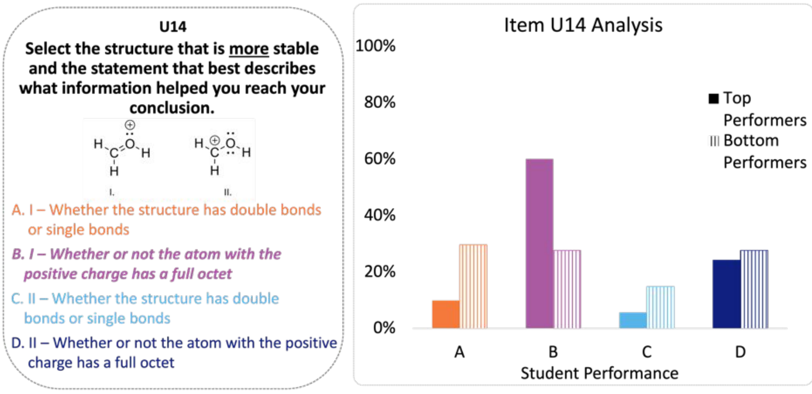
Representational competence should not be assumed. Instructors are not always intentional about developing learner RC (Jones et al., 2022; Linenberger & Holme, 2015; Popova & Jones, 2021). Our data illustrate that instructors should not assume that RC has been acquired after traditional instruction. For example, instructors may assume that *interpreting* the wedge-dash diagram in **Figure 6** should be relatively simple and that learners have developed the fundamental ability to identify implicit hydrogens. However, we administered ORCA at the end of the first semester of organic chemistry, and 30% of participants selected an incorrect answer. This is problematic because the inability to identify all implicit hydrogen atoms may significantly impact learner success in the course. For example, this skill is necessary for identifying chiral centers, completing β-elimination reactions, and solving ^{1}H NMR spectroscopy problems. Our data show that providing learners with continuous practice interpreting implicit atoms in representations is important, even after initial instruction.



Item U15				
Option	Answer	Reasoning	Count	Percent
A	Incorrect	Appropriate	347	23%
B	Incorrect	Inappropriate	157	11%
C	Correct	Appropriate	795	53%
D	Correct	Inappropriate	195	13%

C.

Figure 4. Item U15 (A) with the correct answer in bold and italics. The chart (B) shows the response choice distribution for the top 27% and bottom 27% of participants. The count and percentage of learners selecting each answer are shown in table (C).



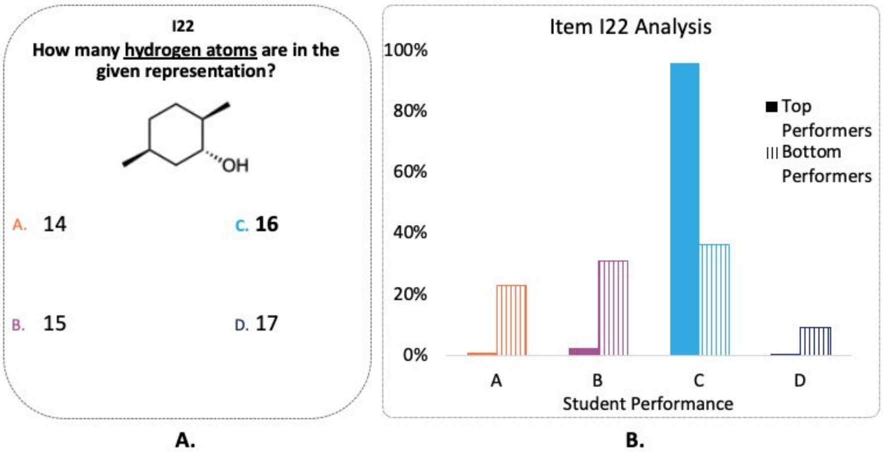
Item U14				
Option	Answer	Reasoning	Count	Percent
A	Correct	Inappropriate	301	20%
B	Correct	Appropriate	610	41%
C	Incorrect	Inappropriate	131	9%
D	Incorrect	Appropriate	452	30%

C.

Please do not adjust margins

ARTICLE

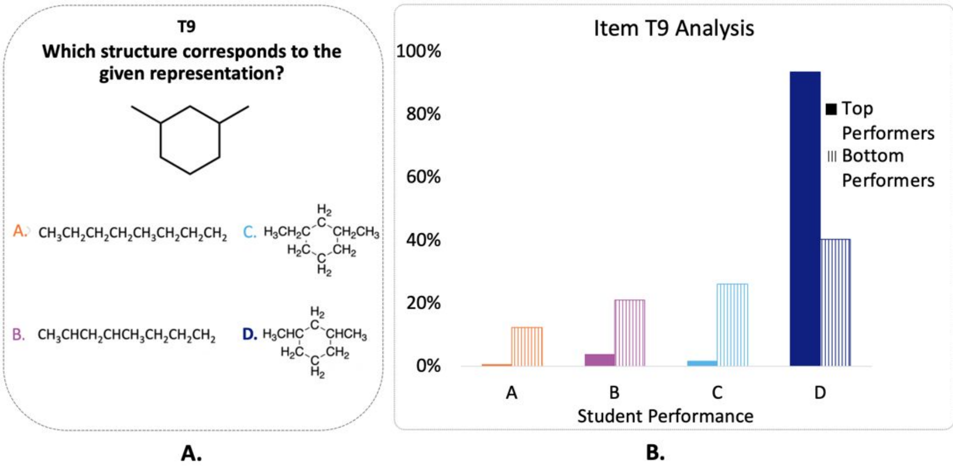
Figure 5. Item U14 (A) with the correct answer in bold and italics. The chart (B) shows the response choice distribution for the top 27% and bottom 27% of participants. The count and percentage of learners selecting each answer are shown in table (C).



Item I22				
Option	Answer	Reasoning	Count	Percent
A	Incorrect	NA	147	10%
B	Incorrect	NA	232	16%
C	Correct	NA	1053	70%
D	Incorrect	NA	62	4%

C.

Figure 6. Item I22 (A) with the correct answer in bold. The chart (B) shows the response choice distribution for the top 27% and bottom 27% of participants. The count and percentage of learners selecting each answer are shown in table (C).



Item T9				
Option	Answer	Reasoning	Count	Percent
A	Incorrect	NA	67	5%
B	Incorrect	NA	171	11%
C	Incorrect	NA	171	11%
D	Correct	NA	1085	73%

C.

Figure 7. Item T9 (A) with the correct answer in bold. The chart (B) shows the response choice distribution for the top 27% and bottom 27% of participants. The count and percentage of learners selecting each answer are shown table (C).

Item T9 is another item that some may assume is easy but had relatively surprising responses. The item requires learners to *translate* between a cyclic skeletal structure and a condensed structure with two of the four answer choices in which the ring is open (**Figure 7A**). Most participants selected the correct answer, but over 15% selected a non-cyclic condensed structure (**Figure 7C**, options A and B). Participants who select these incorrect answer options for a *translate* task do not adequately understand the representations or the chemical phenomena represented, even after a whole semester of organic chemistry. As shown, developing RC should be intentional and repeatedly reinforced throughout the curricula.

Limitations

Several limitations should be acknowledged. First, the ORCA is limited as it includes items related to only three RC skills. However, these skills are among the most commonly taught by chemistry instructors (Jones et al., 2022; Linenberger & Holme, 2015; Popova & Jones, 2021) and reinforced in organic chemistry textbooks (Gurung et al., 2022). More research is needed to investigate and develop quality assessments that will incorporate the remaining RC skills.

Second, the ORCA assesses learner RC in the context of multiple representations of molecular structure but is not exhaustive (e.g., representations such as electrostatic potential maps and ball-and-stick models are not included even though they have unique affordances to students (Farheen et al., 2024)); we selected symbolic representations only to feature representations with similar dimensions and features (Talanquer, 2022). More assessments are needed to capture learner RC with other representations in chemistry.

Third, the *use* items capture learners' ability to make inferences about stability and boiling point only, even though there are other properties that the six target representations could convey. This narrower focus connects to what is covered in instruction and ensures the assessment can be administered in a reasonable timeframe (about 20 minutes). There is potential to consider additional items that focus on other concepts (e.g., aromaticity, acidity).

Fourth, the participants who completed the final assessment also took the assessment during the pilot III administration. There were at least six weeks between assessment administrations (depending on the institution), but this means there is a possibility of practice effects (i.e., improved performance due to familiarity with test items) (American Educational Research Association et al., 2014). Currently, there is no benchmark for when practice effects diminish. Other studies have as low as one month between test administrations (Bretz & Linenberger, 2012).

Lastly, the assessment was administered only online; however, it can be easily adapted to the classroom setting for paper-and-pencil administration. Administering the assessment online was a purposeful choice; this allowed for easy access to learners and a quick collection of data. Assessments given in different conditions (such as virtual versus in class), with differing constraints (*e.g.*, not being able to sketch notes), and scored for correctness versus completion (including for bonus points) all likely have varying but overall consistent data (Harle & Towns, 2011).

Conclusions

Herein, we describe the development of the Organic chemistry Representational Competence Assessment (ORCA). Grounded in the RC framework (Kozma & Russell, 2005), the ORCA captures learner competence with six representations of molecular structure: condensed structures, chair conformations, wedge-dash diagrams, Lewis structures, skeletal structures, and Newman projections. ORCA items include three interconnected RC skills: the ability to *interpret* representations, *translate* between representations, and *use* representations to draw inferences. The assessment is designed to be multiple choice to permit testing (a) with a large number of learners, (b) in a short amount of time, and (c) to allow for quick and easy scoring. Evidence of the validity and reliability of the ORCA data were established through multiple administrations at several institutions in the US and captured via a variety of methods such as response-process interviews, expert panel feedback, CFA, and McDonald's omega (**Table 1**).

Two models were proposed to evaluate the internal structure validity of the data generated: a three-factor Model A and a one-factor Model B. While Model A has a solid theoretical basis, our data showed a better fit with Model B. This single-factor model consolidates the skills into a unified latent RC construct, accounting for the high correlation between the *interpret* and *translate* skill factors in Model A. This model can be readily used to assess organic chemistry learner RC in the context of representations of molecular structure. Our study is an important step toward advancing and reconceptualizing our understanding of RC as a *network of interconnected* skills.

Implications

The ORCA can be used as a formative assessment in instruction or as an assessment instrument in research. The ORCA takes about 20 minutes to complete and can be used by instructors to make the necessary adjustments to their instruction and provide feedback to their students about their competence with representations of molecular structure. A single score from the ORCA can provide a measure of learner RC, as reflected by Model B. Given that formative assessment is regarded as a high-impact instructional practice (National Research Council, 2012; Offerdahl & Arneson, 2019), the use of ORCA to receive and implement formative feedback has the potential to improve learner RC and conceptual understanding.

For example, ORCA can be used as a diagnostic tool early in the course to identify areas where students may need additional support, such as with more complex skills like *using* representations. Based on students' performance, instructors can implement progressive scaffolding, offering explicit guidance early in the semester on how to *use* different representations. As students build confidence, the level of support can be gradually reduced to encourage independent *use* of representations.

At the same time, instructors should not assume that it is sufficient to discuss a given representation once, early in the semester. Our findings show that even by the end of the semester, many students struggled with basic tasks, such as identifying implicit hydrogen atoms in wedge-dash diagrams or distinguishing between cyclic and acyclic molecules. It is, therefore, imperative that instructors periodically incorporate explicit explanations of the features of representations that they use in their instruction without assuming that students mastered these representations at the beginning of the course.

Additionally, it's crucial for students to articulate their reasoning, especially when tackling more complex tasks that require *using* representations to make inferences. Our item analysis revealed that for students to effectively *use* molecular representations to make inferences, they must be able to do four essential steps: (1) interpret the diagrammatic features, (2) understand the underlying concept, (3) identify relevant versus irrelevant features, and (4) connect those relevant features to their conceptual understanding to draw inferences. Without eliciting students' reasoning, an instructor or researcher will not be able to understand which of these steps present the challenge for students when *using* representations to make inferences. This complexity should be taken into consideration when designing assessment tasks or interview prompts.

The ORCA can also be used or adapted as an assessment instrument in studies that aim to advance our understanding of RC as a construct or how to best support learners in developing RC. Prior to using the ORCA, it is imperative that researchers evaluate their data for evidence of validity and reliability (American Educational Research Association et al., 2014; Lazenby et al., 2023). For example, the ORCA can provide evidence of how instructional practices impact learner RC and can be used in longitudinal studies to show how learner RC changes throughout instruction (Kozma et al., 2000). These investigations could help elucidate a learning progression for the development of RC or identify an optimal organic chemistry curriculum for developing RC with representations of molecular structure. While there are existing interventions that support individual RC skills in organic chemistry (Stieff et al., 2016; Stull et al., 2016; Stull & Hegarty, 2016), more research needs to be done to develop interventions that support RC more broadly. In this way, the ORCA can be used as a pre-post measure to provide empirical evidence of approaches that may be useful to support organic chemistry learners.

Future research could also use the ORCA data to conduct Structural Equation Modelling (SEM) to evaluate the relationship between learner RC and other relevant constructs such as their visuospatial ability, conceptual understanding, or success in the organic chemistry course (Nitz et al., 2012; Sim & Daniel, 2014; Stieff et al., 2014, 2018; Stieff & DeSutter, 2020).

While this work focuses on symbolic representations of molecular structure, organic chemistry and other chemistry disciplines also rely on a variety of other representations (e.g., ball-and-stick models, spectra, reaction coordinate diagrams, phase diagrams, and molecular orbital diagrams) that require RC. The process we used to develop the ORCA can be adapted to assess RC skills in these other contexts. Expanding this approach across different contexts will deepen our understanding of how learners develop RC. We encourage the community to build on this work to refine our collective understanding of RC and explore the most effective strategies for supporting learners in developing these important skills.

Ethical Considerations

This study has been approved by the University of North Carolina's Institutional Review Board (IRB #20-0511). IRB guidelines were met at every institution where data were collected and in all stages of the assessment development and evaluation. All participants were informed that their ORCA performance would not negatively impact their course grades. Additionally, the participants could consent or decline consent for their responses to be included in this research study. In accordance with the approved IRB protocol, the findings are presented only in aggregate, and the original data sets cannot be shared due to confidentiality considerations.

Author Contributions

All authors participated in data collection and/or analysis. Author M.P. conceptualized and led the project and obtained funding for the work. F.R. conducted the interviews and assisted in analyzing data. L.W.W. led the qualitative and psychometric analyses. J.R., S.N., and R.K. provided guidance on the psychometric analyses. L.W.W., M.P., and F.R., wrote the manuscript with input from the rest of the authors.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

This work is supported by the National Science Foundation, DUE 2025216. We thank the instructors who administered ORCA, learners who completed ORCA, and experts who evaluated ORCA.

Appendix

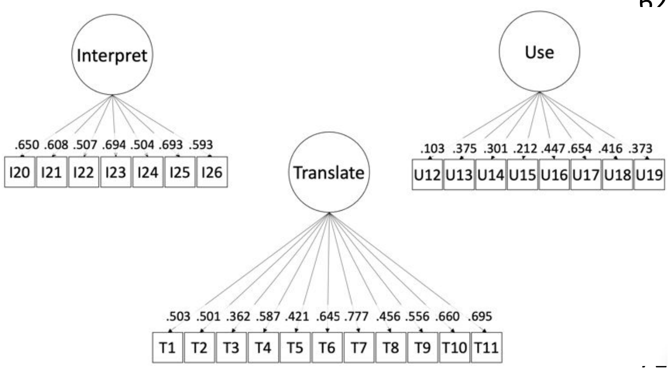


Figure S1. Item loadings of the one-factor models for the *interpret*, *translate*, and *use* scales.

Table S1. Fit statistics and reliability for each latent factor in Model A as a single-factor model.

	Fit Statistics					Reliability
Factor	χ^2	*df*	*p*	RMSEA	CFI	McDonald's Omega (ω)
Interpret	7.709	14	.9039	<.0001	>.999	.807
Translate	200.343	44	<.0001	.049	.950	.843
Use	52.015	20	<.0002	.033	.915	.554

References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. In *Organizational Research Methods* (Vol. 16, Issue 2, pp. 270–301). SAGE Publications Inc.

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198.

Albanese, M. A. (1993). Type K and Other Complex Multiple-Choice Items: An Analysis of Research and Item Properties. *Educational Measurement: Issues and Practice*, *12*(1), 28–33.

American Educational Research Association; American Psychological Association; National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Education Research Association: Washington, DC, 1999.

Arjoon, J. A., Xu, X., & Lewis, J. E. (2013). Understanding the state of the art for measurement in chemistry education research: Examining the psychometric evidence. *Journal of Chemical Education*, *90*(5), 536–545.

Bandalos, D. L. (2018). *Measurement Theory and Applications for the Social Sciences*, New York, USA: The Guilford Press.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, *13*(2), 186–203. 2

Bodner, G. M., & Domin, D. S. (2000). Mental Models: The Role of Representations in Problem Solving in Chemistry. *University Chemistry Education*, *4*(1), 24–30.

Bowen, C. W. (1994). Think-aloud methods in chemistry education: Understanding student thinking. *Journal of Chemical Education*, *71*(3), 184–190.

Bretz, S. L., & Linenberger, K. J. (2012). Development of the enzyme-substrate interactions concept inventory. *Biochemistry and Molecular Biology Education*, *40*(4), 229–233.

Bruice, P. Y. (2016). *Organic Chemistry* (8th ed.). Pearson.

Chang, H. Y. (2018). Students' representational competence with drawing technology across two domains of science. *Science Education*, *102*(5), 1129–1149.

Chang, H. Y., & Tzeng, S. F. (2017). Investigating Taiwanese Students' Visualization Competence of Matter at the Particulate Level. *International Journal of Science and Mathematics Education*, *16*(7), 1207–1226.

Connor, M. C., Finkenstaedt-Quinn, S. A., & Shultz, G. V. (2019). Constraints on organic chemistry students' reasoning during 13C and 1 H NMR spectral interpretation. *Chemistry Education Research and Practice*, *20*(3), 522–541.

Connor, M. C., Glass, B. H., & Shultz, G. V. (2021). Development of the NMR Lexical Representational Competence (NMR-LRC) Instrument As a Formative Assessment of Lexical Ability in1H NMR Spectroscopy. *Journal of Chemical Education*, *98*(9), 2786–2798.

Cooper, M. M., Grove, N., Underwood, S. M., & Klymkowsky, M. W. (2010). Lost in Lewis structures: An investigation of student difficulties in developing representational competence. *Journal of Chemical Education*, *87*(8), 869–874.

Daniel, K. L. (2018). Towards a Framework for Representational Competence in Science Education. In *Towards a Framework for Representational Competence in Science Education* (Vol. 11). Springer International Publishing.

Deng, J. M., Streja, N., & Flynn, A. B. (2021). Response Process Validity Evidence in Chemistry Education Research. *Journal of Chemical Education*, *98*, 3656–3666.

Dickmann, T., Opfermann, M., Dammann, E., Lang, M., & Rumann, S. (2019). What you see is what you learn? The role of visual model comprehension for academic success in chemistry. *Chemistry Education Research and Practice*, *20*(4), 804–820.

diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and Instruction*, *22*(3), 293–331.

diSessa, Andrea, A., & Sherin, B. L. (2000). Meta-representation: an introduction. *Journal of Mathematical Behavior*, *19*, 385–398.

Farheen, A., & Lewis, S. E. (2021). The impact of representations of chemical bonding on students' predictions of chemical properties. *Chemistry Education Research and Practice*, *22*(4), 1035–1053.

Farheen, A., Martin, N., & Lewis, S. E. (2024). Student perceptions of partial charges and nucleophilicity/electrophilicity when provided with either a bond-line, ball-and-stick, or electrostatic potential map for molecular representation. *Chemistry Education Research and Practice*, *25*(1), 343–359.

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 439–492). IAP Information Agre Publishing.

Gilbert, J. K. (2005). Visualization: A Metacognitive Skill in Science and Science Education. In *Visualization in Science Education* (pp. 9–27). Springer.

Glaser, B. G. (1965). *The Constant Comparative Method of Qualitative Analysis*. 12, 436–445.

Gurung, E., Jacob, R., Bunch, Z., Thompson, B., & Popova, M. (2022). Evaluating the Effectiveness of Organic Chemistry Textbooks for Promoting Representational Competence. *Journal of Chemical Education*, *99*, 2044–2054.

Halverson, K. L., & Friedrichsen, P. (2013). Learning tree thinking: Developing a new framework of representational competence. In *Multiple Representations in Biological Education* (Vol. 7, pp. 185–201).

Hancock, G., & An, J. (2018). *ITEMS Module Companion Article: Scale Reliability in Structural Equation Models*. 1–16.

Harle, M., & Towns, M. (2011). A review of spatial ability literature, its connection to chemistry, and implications for instruction. *Journal of Chemical Education*, *88*(3), 351–360.

Herrington, D. G., & Daubenmire, P. L. (2014). Using interviews in CER projects: Options, considerations, and limitations. In D. M. Bunce & R. S. Cole (Eds.), *Tools of Chemistry Education Research (ACS Symposium Series)* (Vol. 1166, pp. 31–59).

Hinze, S. R., Rapp, D. N., Williamson, V. M., Shultz, M. J., Deslongchamps, G., & Williamson, K. C. (2013). Beyond ball-and-stick: Students' processing of novel STEM visualizations. *Learning and Instruction*, *26*, 12–21.

Hiong, S. J., & Daniel, E. G. S. (2015). Same Substances, but They "See" Different Things: Student-Generated Representations as a Research Tool. *Journal of Science and Mathematics Education in Southeast Asia*, *38*(1), 1–31.

Höst, G. E., Schönborn, K. J., & Palmerius, K. E. L. (2012). Students use of three different visual representations to interpret whether molecules are polar or nonpolar. *Journal of Chemical Education*, *89*(12), 1499–1505.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.

Johnstone, A. H. (1993). The development of chemistry teaching: A changing response to changing demand. *Journal of Chemical Education*, *70*(9), 701–705.

Jones, T., Romanov, A., Pratt, J. M., & Popova, M. (2022). Multi-framework case study characterizing organic chemistry instructors' approaches toward teaching about representations. *Chemistry Education Research and Practice*, *23*(4), 930–947.

Kohl, P. B., & Finkelstein, N. D. (2006). Effects of representation on students solving physics problems: A fine-grained characterization. *Physical Review Special Topics - Physics Education Research*, *2*(1), 1–12.

Komperda, R., Pentecost, T. C., & Barbera, J. (2018). Moving beyond Alpha: A Primer on Alternative Sources of Single-Administration Reliability Evidence for Quantitative Chemistry Education Research. *Journal of Chemical Education*, *95*(9), 1477–1491.

Kozma, R. B., & Russell, J. (1997). Multimedia and Understanding: Expert and Novice Responses to Different Representations of Chemical Phenomena. *Journal of Research in Science Teaching*, *34*(9), 949–968.

Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The Roles of Representations and Tools in the Chemistry Laboratory and Their Implications for Chemistry Learning. *Journal of the Learning Sciences*, *9*(2), 105–143.

Kozma, & Russell, J. (2005). Students Becoming Chemists: Developing Representational Competence. In J. K. Gilbert (Ed.), *Visualization in Science Education* (pp. 121–145). Springer.

Lansangan, R., Orleans, A., & Camacho, V. M. (2007). Representational competence of secondary chemistry students in understanding selected chemical principles. *International Journal of Science Education*, *24*(1), 7930–7934.

Lansangan, R., Orleans, A., & Camacho, V. M. (2018). Assessing Conceptual Understanding in Chemistry Using Representation. *Advanced Science Letters*, *24*(11), 7930–7934.

Lave, J. (1991). Situated learning in community of practice. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 63–82). American Psychological Association.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Lazenby, K., Tenney, K., Marcroft, T. A., & Komperda, R. (2023). Practices in instrument use and development in chemistry education research and practice 2010-2021. *Chemistry Education Research and Practice*, *24*(3), 882–895. https://doi.org/10.1039/d2rp00275b

Linenberger, K. J., & Holme, T. A. (2015). Biochemistry instructors' views toward developing and assessing visual literacy in their courses. *Journal of Chemical Education*, *92*(1), 23–31.

Malkewitz, C. P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α, McDonald's ω and the greatest lower bound. *Social Sciences & Humanities Open*, *7*(1), 100368.

Maroo, J. D., & Johnson, S. L. (2018). The Use of a Representational Triplet Model as the Basis for the Evaluation of Students' Representational Competence. In K. Daniel (Ed.), *Towards a Framework for Representational Competence in Science Education* (pp. 247–262). Springer.

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: An example of the two-way street between cognition and instruction. *New Directions for Teaching and Learning*, *89*, 55–71. https://doi.org/10.1002/tl.47

McNeish, D., An, J., & Hancock, G. R. (2018). The Thorny Relation Between Measurement Quality and Fit Index Cutoffs in Latent Variable Models. *Journal of Personality Assessment*, *100*(1), 43–52.

McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for one-factor models. *Behavior Research Methods*, *55*(3), 1157–1174.

Miller, K., & Kim, T. (2017). Examining student heuristic usage in a hydrogen bonding assessment. *Biochemistry and Molecular Biology Education*, *45*(5), 411–416.

Murphy, K., Holme, T., Zenisky, A., Caruthers, H., & Knaus, K. (2012). Building the ACS exams anchoring concept content map for undergraduate chemistry. *Journal of Chemical Education*, *89*(6), 715–720.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide.* (Eighth Edition). Muthén & Muthén.

National Research Council. (2012a). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2012b). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering.* S.R. Singer, N.R. Nielse, and H.A. Schweingruber, Editors. Committee on the Status, Contributions, and Future Directions of Discipline-Based Education Research. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nitz, S., Nerdel, C., & Prechtl, H. (2012). Modeling the Relationship between Representational Competence and Domain Knowledge. In *EARLI Conference SIGc* (pp. 160–162).

Offerdahl, E. G., & Arneson, J. B. (2019). Formative Assessment to Improve Student Learning in Biochemistry [Chapter]. In *ACS Symposium Series: Biochemistry Education: From Theory to Practice* (Vol. 1337, pp. 197–218). American Chemical Society.

Padalkar, S., & Hegarty, M. (2012). Improving Representational Competence in Chemistry with Model-Based Feedback. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*(34), 2162–2167.

Padalkar, S., & Hegarty, M. (2015). Models as feedback: Developing representational competence in chemistry. *Journal of Educational Psychology*, *107*(2), 451–466.

Pande, P., & Chandrasekharan, S. (2022). Expertise as Sensorimotor Tuning: Perceptual Navigation Patterns Mark Representational Competence in Science. *Research in Science Education*, *52*(2), 725–747.

Pande, P., Shah, P., & Chandrasekharan, S. (2015). How do experts and novices navigate chemistry representations-an eye-tracking investigation. *Proceedings of EPISTEME-6*, 102–109.

Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods* (3rd ed.). Sage Publications.

Popova, M., & Bretz, S. L. (2018a). “It’s Only the Major Product That We Care About in Organic Chemistry”: An Analysis of Students’ Annotations of Reaction Coordinate Diagrams. *Journal of Chemical Education*, *95*(7), 1086–1093.

Popova, M., & Bretz, S. L. (2018b). Organic chemistry students’ challenges with coherence formation between reactions and reaction coordinate diagrams. *Chemistry Education Research and Practice*, *19*(3), 732–745.

Popova, M., & Bretz, S. L. (2018c). Organic chemistry students’ interpretations of the surface features of reaction coordinate diagrams. *Chemistry Education Research and Practice*, *19*(3), 919–931.

Popova, M., & Jones, T. (2021). Chemistry instructors’ intentions toward developing, teaching, and assessing student representational competence skills. *Chemistry Education Research and Practice*, *2*(3), 733–748.

Raker, J. R., & Holme, T. A. (2013). A historical analysis of the curriculum of organic chemistry using ACS exams as artifacts. *Journal of Chemical Education*, *90*(11), 1437–1442.

Rosseel, Y. (2012). *Journal of Statistical Software lavaan: An R Package for Structural Equation Modeling*. http://www.jstatsoft.org/

Rotich, F., Ward, L., Beck, C., & Popova, M. (2024). Attention is currency: how surface features of Lewis structures influence organic chemistry student reasoning about stability. *Chemistry Education Research and Practice*. Advanced Article.

Saldaña, J., (2013) The coding manual for qualitative researchers. Second edn. Los Angeles: SAGE Publications.

Schönborn, K. J., & Anderson, T. R. (2008). A Model of Factors Determining Students’ Ability to Interpret External Representations in Biochemistry. *International Journal of Science Education*, *31*(2), 193–232.

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, *13*(2), 227–237.

Sim, J. H., & Daniel, E. G. S. (2014). Representational competence in chemistry: A comparison between students with different levels of understanding of basic chemical concepts and chemical representations. *Cogent Education*, *1*(1).

Stieff, M., & DeSutter, D. (2020). Sketching, not representational competence, predicts improved science learning. *Journal of Research in Science Teaching*, 1–29.

Stieff, M., Hegarty, M., & Deslongchamps, G. (2011). Identifying representational competence with multi-representational displays. *Cognition and Instruction*, *29*(1), 123–145.

Stieff, M., Lira, M., & Desutter, D. (2014). Representational Competence and Spatial Thinking in STEM Spatial Thinking in STEM. *Proceedings of International Conference on the Learning Sciences, ICLS,* 2, 987-991.

Stieff, M., Origenes, A., DeSutter, D., Lira, M., Banevicius, L., Tabang, D., & Cabel, G. (2018). Operational constraints on the mental rotation of STEM representations. *Journal of Educational Psychology*, *110*(8), 1160–1174.

Stieff, M., Scopelitis, S., Lira, M. E., & Desutter, D. (2016). Improving Representational Competence with Concrete Models. *Science Education*, *100*(2), 344–363.

Stull, A. T., Gainer, M., Padalkar, S., & Hegarty, M. (2016). Promoting Representational Competence with Molecular Models in Organic Chemistry. *Journal of Chemical Education*, *93*(6), 994–1001.

Stull, A. T., & Hegarty, M. (2016). Model manipulation and learning: Fostering representational competence with virtual and concrete models. *Journal of Educational Psychology*, *108*(4), 509–527.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296.

Talanquer, V. (2014). Chemistry Education: Ten heuristics to tame. *Journal of Chemical Education*, *91*(8), 1091–1097.

Talanquer, V. (2022). The Complexity of Reasoning about and with Chemical Representations. *JACS Au*, *2*(12), 2658–2669.

Towns, M. H. (2008). Mixed Methods Designs in Chemical Education Research. In D. Bunce & R. Cole (Eds.), *Nuts and Bolts of Chemistry Education Research* (Vol. 15, pp. 135–148). American Chemical Society.

Towns, M. H. (2014). Guide to developing high-quality, reliable, and valid multiple-choice assessments. *Journal of Chemical Education*, *91*(9), 1426–1431.

Vlacholia, M., Vosniadou, S., Roussos, P., Salta, K., Kazi, S., Sigalas, M., & Tzougraki, C. (2017). Changes in visual/spatial and analytic strategy use in organic chemistry with the development of expertise. *Chemistry Education Research and Practice*, *18*(4), 763–773.

Wang, Z., Chi, S., Luo, M., Yang, Y., & Huang, M. (2017). Development of an instrument to evaluate high school students' chemical symbol representation abilities. *Chemistry Education Research and Practice*, *18*(4), 875–892.

Ward, L. W., Rotich, F., Hoang, J., & Popova, M. (2022). Representational Competence Under the Magnifying Glass—The Interplay Between Student Reasoning Skills, Conceptual Understanding, and the Nature of Representations. In N. Graulich & G. Shultz (Eds.), *Student Reasoning in Organic Chemistry: Research Advances and Evidence-based Instructional Practices* (pp. 36–55). Royal Society of Chemistry.

Wright, L., & Oliver-Hoyo, M. T. (2020). Student assumptions and mental models encountered in IR spectroscopy instruction. *Chemistry Education Research and Practice*, *21*(1), 426–437.

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409–428.

Electronic Supplemental Information
Design, development, and evaluation of the Organic chemistry Representational Competence Assessment (ORCA)

Lyniesha Ward,[a] Fridah Rotich,[a] Jeffrey R. Raker,[b] Regis Komperda,[c] Sachin Nedungadi,[d] and Maia Popova*[a]

[a]University of North Carolina at Greensboro, Greensboro, North Carolina, USA
[b]University of South Florida, Tampa, Florida, USA
[c]San Diego State University, San Diego, California, USA
[d]University of Nebraska at Omaha, Omaha, Nebraska, USA
*E-mail: m_popova@uncg.edu

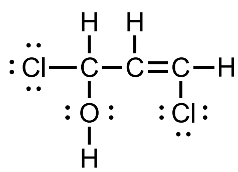
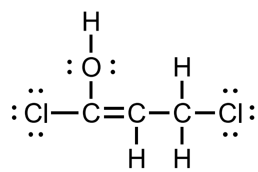
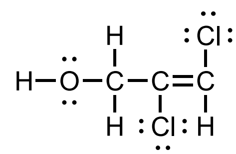
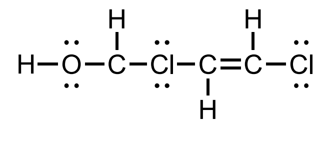
NOTE: If you are interested in using ORCA, please contact the corresponding author.

To Page 1

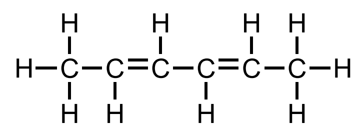
Directions: Identify the choice that best completes the statement or answers the question.

1. Which structure corresponds to the given representation?

HOCHClCH=CHCl

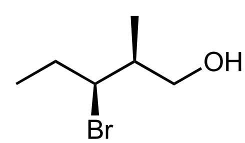
a.  b.  c.  d. 

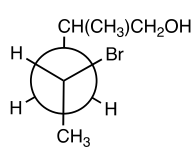
2. Which structure corresponds to the given representation?

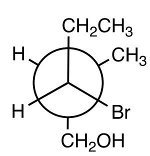


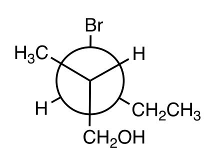
a. $CH_3CH{=}CHCH_2CH_3$ b. $CH_3(CH_2)_2CH_3$ c. $CH_3(CH_2)_4CH_3$ d. $CH_3(CH{=}CH)_2CH_3$

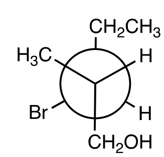
3. Which structure is <u>different</u> from the given representation?



a. 

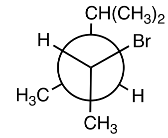
b. 

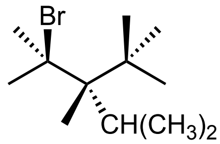
c. 

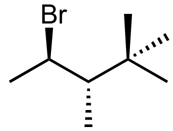
d. 

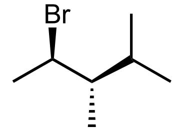
To Page 1

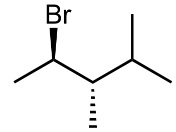
4. Which structure corresponds to the given representation?



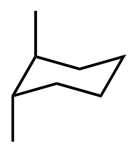
a. 

b. 

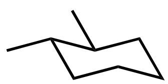
c. 

d. 

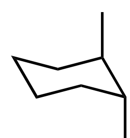
5. Use the given representation below to answer this question.



Select the structure that is the same molecule as the given representation and the statement that best describes what information you focused on to reach your conclusion.



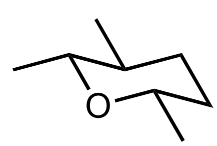
I.

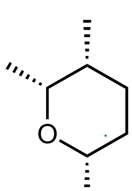


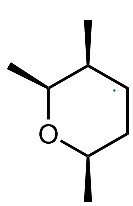
II.

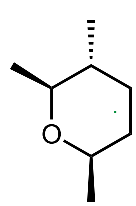
a. I – Whether the groups are on the left side in relation to the ring
b. I – Whether the groups are up versus down in relation to the ring
c. II – Whether the groups are axial versus equatorial in the ring
d. II – Whether the groups are on the outer edges of the ring

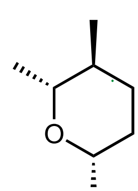
6. Which structure corresponds to the given representation?



a. 

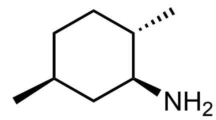
b. 

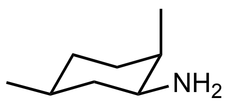
c. 

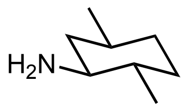
d. 

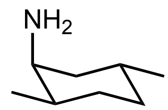
To Page 1

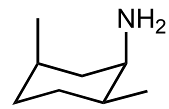
7. Which structure corresponds to the given representation?



a. 

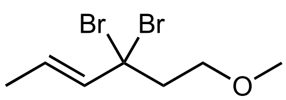
b. 

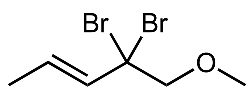
c. 

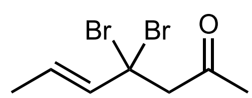
d. 

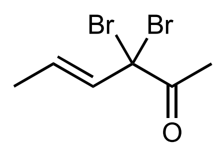
8. Which structure corresponds to the given representation?

$CH_3CH{=}CHCBr_2CH_2COCH_3$

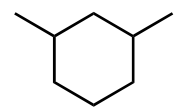
a. 

b. 

c. 

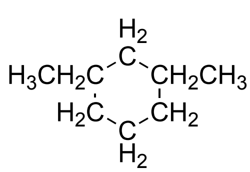
d. 

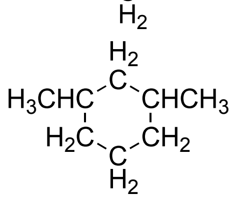
9. Which structure corresponds to the given representation?



a. $CH_3CH_2CH_2CH_2CH_3CH_2CH_2CH_2$

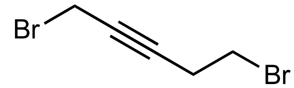
b. $CH_3CHCH_2CHCH_3CH_2CH_2CH_2$

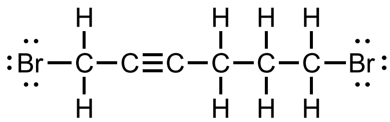
c. 

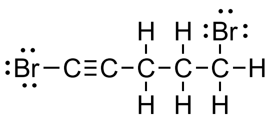
d. 

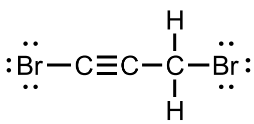
To Page 1

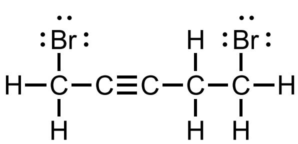
10. Which structure corresponds to the given representation?



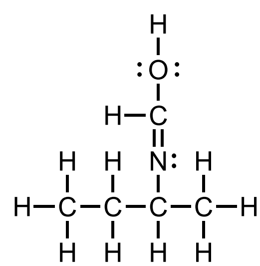
a. 

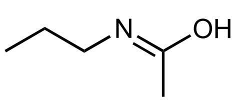
b. 

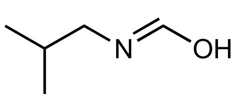
c. 

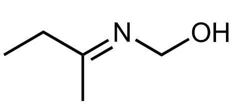
d. 

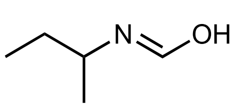
11. Which structure corresponds to the given representation?



a. 

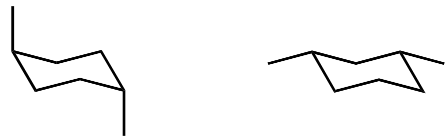
b. 

c. 

d. 

To Page 1

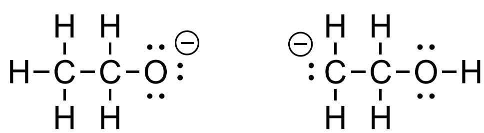
12. Select the structure that is the <u>more</u> stable and the statement that best describes what information you focused on to reach your conclusion.



I. II.

a. I – The amount of space between the groups
b. I – Whether the groups are pointing up or down
c. II – The amount of space between the groups
d. II – Whether the groups are pointing up or down

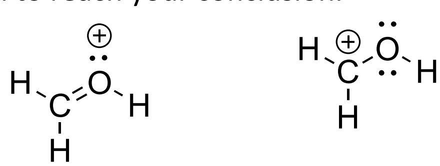
13. Select the structure that is the <u>less</u> stable and the statement that best describes what information you focused on to reach your conclusion.



I. II.

a. I – The electronegativity of the atom with the negative charge
b. I – The absence of a full octet on the atom with the negative charge
c. II – The electronegativity of the atom with the negative charge
d. II – The absence of a full octet on the atom with the negative charge

14. Select the structure that is the <u>more</u> stable and the statement that best describes what information you focused on to reach your conclusion.

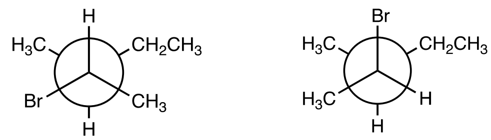


I. II.

a. I – Whether the structure has double bonds or single bonds
b. I – Whether or not the atom with the positive charge has a full octet
c. II – Whether the structure has double bonds or single bonds
d. II – Whether or not the atom with the positive charge has a full octet

To Page 1

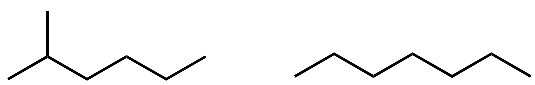
15. Select the structure that is the <u>less</u> stable and the statement that best describes what information you focused on to reach your conclusion.



I. II.

a. I – The size and proximity of the groups to one another
b. I – Whether or not electronegative groups are near other groups
c. II – The size and proximity of the groups to one another
d. II – Whether or not electronegative groups are near other groups

16. Select the structure with a <u>lower</u> boiling point and the statement that best describes the information you focused on to reach your conclusion.



I. II.

a. I – Whether the molecule has branching
b. I – Whether the molecule is symmetrical or asymmetrical
c. II – Whether the molecule has branching
d. II – Whether the molecule is symmetrical or asymmetrical

17. Which of the following compounds has the <u>highest</u> boiling point?

a. $CH_3CH_2CH_2CH_2OH$

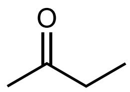
b. $CH_3CH_2CH_2CHO$

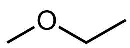
c. $CH_3CH_2OCH_2CH_3$

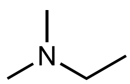
To Page 1

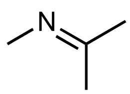
d. $CH_3COCH_2CH_3$

18. Which of the following compounds has the <u>highest</u> boiling point?

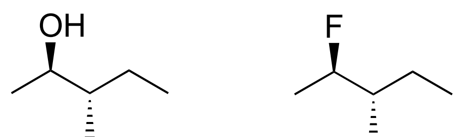
a. 

b. 

c. 

d. 

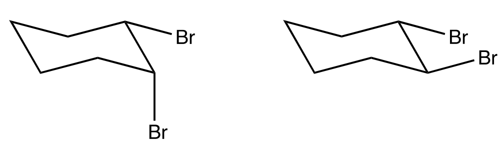
19. Select the structure with a <u>higher</u> boiling point and the statement that best describes the information you focused on to reach your conclusion.



I. II.

a. I – How strong the bonds are between atoms
b. I – How strong the interactions are between molecules
c. II – How strong the bonds are between atoms
d. II – How strong the interactions are between molecules

20. Select the structure that has bromine atoms on opposite sides in relation to the ring and the statement that best describes what information you focused on to reach your conclusion.

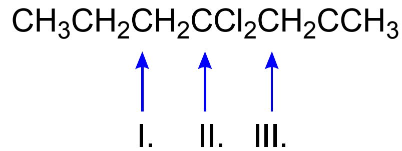


I. II.

a. I – Whether the bromine atoms are axial versus equatorial in relation to the ring
b. I – Whether the bromine atoms are up versus down in relation to the ring
c. II - Whether the bromine atoms are axial versus equatorial in relation to the ring
d. II - Whether the bromine atoms are up versus down in relation to the ring

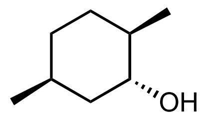
To Page 1

21. To which labeled atom(s) is/are the <u>chlorine atoms</u> connected in the given representation?



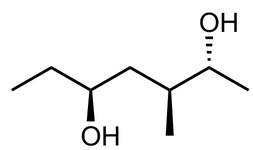
a. Carbon I
b. Carbon II
c. Carbon III
d. Carbons II, III

22. How many <u>hydrogen</u> atoms are in the given representation?



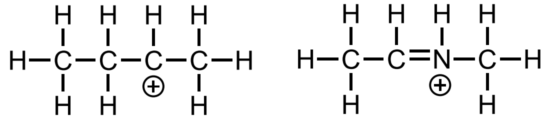
a. 14
b. 15
c. 16
d. 17

23. Which best describes the given representation?



a. The broken dash line is directed out of the page.
b. The bold wedge lines are directed into the page.
c. The thin lines are flat in the plane of the page. The bold wedge and broken dash lines are not.
d. Neither the thin lines, the bold wedge lines, nor the broken dash lines are flat on the plane of the page.

24. Select the structure(s) that has/have an atom with an incomplete octet and the statement that best describes what information you focused on to reach your conclusion.

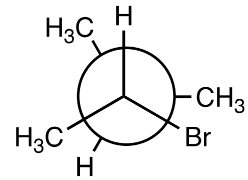


I. II.

a. I – It is missing a pair of electrons.
b. II – It is missing a pair of electrons.
c. I, II – They have a positive formal charge.
d. I, II – Their outer shells are not filled.

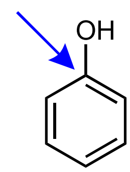
To Page 1

25. How many carbon atoms are in the given representation?



a. 3
b. 4
c. 5
d. 6

26. Which best describes the atom indicated with an arrow in the given representation?



It is connected to…

a. Two carbons and one oxygen.
b. Three carbons and one oxygen.
c. Two carbons, one oxygen, and one hydrogen.
d. Three carbons, one oxygen, and one hydrogen.

To Page 1

Data Availability Statement

Design, development, and evaluation of the Organic chemistry Representational Competence Assessment (ORCA)

Lyniesha Ward,[a] Fridah Rotich,[a] Jeffrey R. Raker,[b] Regis Komperda,[c] Sachin Nedungadi,[d] and Maia Popova*[a]

[a]University of North Carolina at Greensboro, Greensboro, North Carolina, USA
[b]University of South Florida, Tampa, Florida, USA
[c]San Diego State University, San Diego, California, USA
[d]University of Nebraska at Omaha, Omaha, Nebraska, USA
*E-mail: m_popova@uncg.edu

In agreement with the confidentiality measures and procedures approved by the University of North Carolina's Institutional Review Board (IRB # 20-0511), the data cannot be made available.