Research article

Sachin Nedungadi, Michael D. Mosher, Sue Hyeon Paek, Richard M. Hyslop and Corina E. Brown*

Development and psychometric analysis of an inventory of fundamental concepts for understanding organic reaction mechanisms

<https://doi.org/10.1515/cti-2021-0009>

Received March 19, 2021; accepted May 17, 2021; published online June 10, 2021

Abstract: The fundamental concepts for organic reaction mechanisms inventory (FC-ORMI) is a multiple-choice instrument designed to assess students' conception of fundamental concepts for understanding organic reaction mechanisms. The concepts were identified from open-ended interviews and a national survey of organic chemistry instructors reported in a previous study. This manuscript describes the development of the inventory items related to these identified concepts and the psychometric analysis of the instrument. In the developmental stage, open-ended questions were administered to first-semester organic chemistry students ($N = 138$), and open-ended interviews were conducted with students ($N = 22$) from the same pool to gain insight into their thought processes. The answers revealed alternate conceptions which were used to formulate distractors for the inventory. A pilot version and a beta version of the inventory were administered to 105 and 359 first-semester organic chemistry students, respectively. From these administrations, the 26-item alpha version was developed and administered to first-semester undergraduate organic chemistry students ($N = 753$). Psychometric analysis was conducted at the item and test level using Classical Test Theory and Rasch analysis. The results indicate that the items on the FC-ORMI function well to reveal students' alternate conceptions. The instrument meets the acceptable standards of validity and reliability for concept inventories.

Keywords: fundamental concepts; organic chemistry education; organic reaction mechanisms; Rasch analysis.

Introduction

The difficulties students face with understanding organic reaction mechanisms have been a source of interest in chemical education research. Studies have focused on understanding how students use their organic chemistry knowledge to solve mechanistic problems (Anderson & Bodner, 2008; Bhattacharyya & Bodner, 2005; Bhattacharyya, 2013, 2014; Ferguson & Bodner, 2008; Grove, Cooper, & Cox, 2012; Grove, Cooper, & Rush, 2012; Kraft, Strickland, & Bhattacharyya, 2010). However, no work has been reported on the development of a diagnostic tool in the form of a concept inventory for the large-scale assessment of students' understanding of fundamental concepts for understanding organic reaction mechanisms. Such a tool could

*Corresponding author: Corina E. Brown, Department of Chemistry and Biochemistry, University of Northern Colorado, Greeley, CO, USA, E-mail: corina.brown@unco.edu

Sachin Nedungadi, Department of Chemistry and Biochemistry, University of Northern Colorado, Greeley, CO, USA; and Department of Chemistry and Biochemistry, University of Wisconsin Milwaukee, Milwaukee, WI, USA

Michael D. Mosher and Richard M. Hyslop, Department of Chemistry and Biochemistry, University of Northern Colorado, Greeley, CO, USA

Sue Hyeon Paek, Department of Psychological Sciences, University of Northern Colorado, Greeley, CO, USA

assist instructors in identifying alternate conceptions that students have developed. The purpose of this study was the development and psychometric analysis of the Fundamental Concepts for Organic Reaction Mechanisms Inventory (FC-ORMI) to assess students' conception of fundamental concepts needed for understanding organic reaction mechanisms.

Concept inventories are usually multiple-choice assessments that use identified student alternate conceptions as distractors (Richardson, 2004). There have been several concept inventories developed in chemistry as diagnostic tools to determine students' alternate conceptions, and multiple methodologies have been utilized for developing suitable distractors (Atkinson, Popova, Croisant, Reed, & Bretz 2020; Brandriet & Bretz, 2014; Bretz & Linenberger, 2012; Bretz & Murata Mayo, 2018; Brown, Hyslop, & Barbera, 2015; Chu, Treagust, Yeo, & Zadnik, 2012; Dick-Perez, Luxford, Windus, & Holme, 2016; Luxford & Bretz, 2014; McClary & Bretz, 2012; Mulford & Robinson, 2002; Othman, Treagust, & Chandrasegaran, 2008; Villafañe, Bailey, Loertscher, Minderhout, & Lewis, 2011; Wren & Barbera, 2013). Many of these inventories have a two-tier design where the questions are in linked pairs with an answer tier followed by a response tier (Treagust, 1988). These inventories have been extensively developed in different areas of chemistry; however, with the exception of the inventory developed to assess organic chemistry students' understanding of acids and bases (McClary & Bretz, 2012) and the inventory developed to assess organic chemistry students' understanding of reaction coordinate diagrams (Atkinson et al., 2020), there are no concept inventories developed and reported in the literature to assess students' conception of other fundamental concepts needed for understanding organic reaction mechanisms.

An important step in the development of concept inventories is establishing the validity and reliability of the data obtained using the instrument (Bandalos, 2018). Classical Test Theory (CTT) is routinely used as the first step in exploring item functioning because it involves simple mathematical techniques, and it requires moderate sample sizes (Hambleton & Jones, 1993). The majority of the concept inventories developed in chemical education utilize CTT to establish validity and reliability of the data obtained (Brandriet & Bretz, 2014; Bretz & Linenberger, 2012; Bretz & Murata Mayo, 2018; Brown et al., 2015; Chu et al., 2012; Dick-Perez et al., 2016; Luxford & Bretz, 2014). The development of a concept inventory is an iterative process, and CTT provides a quick method to check the functioning of items on a smaller sample which helps justify the modification or elimination of items during the developmental phases.

In CTT, item and person parameters are sample dependent; however, in item response theory (IRT) if the model fits the data, then item and person parameters are sample independent. IRT models, such as the Rasch model, generally require larger sample sizes. Rasch analysis helps researchers determine the probability of an examinee answering an item which gives details about item quality and person ability beyond just the test scores (Hambleton & Jones, 1993).

The utility of Rasch analysis for psychometric assessment and as an analysis technique in science education has been reported (Boone & Scantlebury, 2006; Boone, Townsend, & Staver, 2011). Other than a few exceptions (Hadenfeldt, Bernholt, Liu, Neumann, & Parchmann, 2013; Nedungadi, Paek, & Brown, 2019; Wei, Liu, Wang, & Wang, 2012; Wren & Barbera, 2014), Rasch analysis has not been used frequently when designing concept inventories in chemical education. However, it has been used to add to psychometric data on established concept inventories (Barbera, 2013; Pentecost & Barbera, 2013; Taskin, Bernholt, & Parchmann, 2015). In chemical education, researchers tend to favor reporting particular types of evidence such as test-retest reliability and internal consistency over response process validity while developing concept inventories (Arjoon, Xu, & Lewis, 2013). There is a need for users of these instruments to have a more detailed report on the quality of the assessment tool to use the instrument effectively.

When developing concept inventories it is important to establish proper functioning of the instrument by conducting a rigorous psychometric analysis of the instrument. Rasch analysis together with CTT can provide adequate psychometric data thereby establishing the robustness of the FC-ORMI as an assessment instrument. The present study was aimed at the design and development of the FC-ORMI, and psychometric assessment using CTT and Rasch analysis. The reliability and validity of the data obtained using the FC-ORMI were also established.

Methodology

Protection of human subjects

Institutional Review Board approval was obtained through the university where the research was conducted before data collection for all phases of the study. All interviews were digitally recorded and all written answers to interview questions were recorded using a digital pen and paper (Livescribe, 2016). The participants in the open-ended interviews were given a code (SS# for student subject followed by a number) to maintain their anonymity.

Study design

This research was conducted in various steps. Identification of pertinent concepts, the inventory development, test administration, and psychometric analysis constitute the steps in this study. Figure 1 gives the general model for the study design.

Step 1: Identification of Pertinent Concepts

The fundamental concepts for understanding organic reaction mechanisms were obtained by first conducting interviews with organic chemistry instructors ($N = 11$) and then generalizing the results by conducting a national survey of organic chemistry instructors ($N = 183$). The results yielded 10 fundamental concepts for understanding organic reaction mechanisms based on the opinion of organic chemistry instructors (Nedungadi & Brown, 2020).

Step 2: Inventory Development

Questions representing each of the 10 concepts (see Supplementary Material) were administered to 138 first-semester undergraduate organic chemistry students at a D/PU university (doctoral/professional university) in the Rocky Mountain region of the U.S.A. There were approximately three questions representing each concept.

The questions were administered in an open-ended format and asked for an answer and explanation which contributed to the construction of the two tiers of the inventory. The responses to the questions were analyzed for common alternate conceptions. Some of the most commonly occurring alternate conceptions were subsequently used as distractors for the multiple-choice format of the questions.

Open-ended interviews were also conducted with first-semester undergraduate organic chemistry students. The guiding questions are presented in the Supplementary Material. A convenient sample of 22 students volunteered to participate. Students were presented with the items in both open-ended formats and multiple-choice formats. The interviews gave information on students' thought processes and their use of conceptual knowledge while answering the questions which further assisted in the development of the items.

Figure 1: General model for study design.

Step 3: Test Administration

The pilot version of the inventory containing 50 items (see Supplementary Material) was administered to 105 first-semester undergraduate organic chemistry students at a D/PU university in the Rocky Mountain region of the U.S.A. The beta version of the inventory containing 27 items (see Supplementary Material) was administered to 359 first-semester undergraduate organic chemistry students of which 264 participants were from two different R1 universities (doctoral university: very high research activity) and 95 participants were from an M1 university (masters colleges and universities: larger programs) in the Rocky Mountain region of the U.S.A.

The alpha version of the inventory containing 26 items (see Supplementary Material) was administered to 753 first-semester undergraduate organic chemistry students at 14 different universities throughout the U.S.A. The list of schools and the number of students from each school are shown in Table 1.

The only criterion for participant selection was that the participants had to be enrolled in a first-semester organic chemistry class. The test was administered to the participants one month into the semester before they started the study of reaction mechanisms. The participants were provided with a periodic table, and they were not allowed to use any other instructional materials during the test. Most of the items in all versions of the FC-ORMI were based on a two-tier design with an answer tier followed by a response tier (Treagust, 1988). There was no time limit, but the participants completed the test on a bubble sheet in approximately 45 min for the pilot version and 25 min for both the beta and alpha versions. The tests were administered in person and participants took it during their regular class period.

Step 4: Psychometric Analysis

The psychometric analysis of the FC-ORMI at the item level was conducted using CTT and Rasch analysis to obtain information regarding the functioning of the items. Item difficulty, item discrimination, and distractor analysis in CTT were used. Item difficulty (P) represents the proportion of participants responding correctly (Doran, 1980). Items that have difficulty values between 0.3 and 0.8 for a criterion-referenced test are said to be acceptable with very easy items being above 0.8 and very difficult items being below 0.3 (Kaplan & Saccuzzo, 1997). Item discrimination indices (D) are measures of the degree to which an item distinguishes between the high performers and low performers. For larger samples ($N > 200$), the high performing and low performing groups for calculating discrimination indices are created from the top and bottom 27% of respondents and for smaller samples from the top and bottom 50% of respondents (Kelley, 1939). Discrimination values greater than or equal to 0.3 are deemed acceptable and those below 0.3 may require further investigation (Ebel & Frisbie, 2006).

In Rasch analysis, fit statistics were used to analyze item fit to the Rasch model, and a Wright map was used to compare item difficulty and person ability. The responses were scored as correct or incorrect and aggregated to yield the total score for item analysis in CTT. For Rasch analysis one assumption is unidimensionality and to satisfy this assumption, local independence must be met. Local independence of the items suggests that how a participant answers one item is not dependent on how the participant answers another item (Bond & Fox, 2015). Therefore, for Rasch analysis, the items that are linked pairs were scored as 1 if both tiers were answered correctly and 0 if one or both of the tiers were answered incorrectly.

The psychometric analysis at the test level was conducted by assessing the validity and reliability of the data obtained using the FC-ORMI. The validity of the data obtained using the instrument was established at various stages during the development process using construct validity and employed the subjective measures of face validity and content validity. Additionally, the

Table 1: List of schools and number of students for alpha administration.

School	Region in the U.S.	Carnegie classification	Number of students
A	West North Central	Baccalaureate Colleges: diverse fields	21
B	Mountain	M1: Master's Colleges and Universities – larger programs	53
C	Mountain	R1: Doctoral Universities – Very high research activity	70
D	Mountain	D/PU: Doctoral/Professional Universities	121
E	Mountain	Associate's Colleges: high transfer-mixed traditional/non-traditional	44
F	South Atlantic	D/PU: Doctoral/Professional Universities	42
G	Mountain	R2: Doctoral Universities – high research activity	161
H	Middle Atlantic	Baccalaureate Colleges: arts and science focus	13
I	East North Central	Baccalaureate Colleges: arts and science focus	31
J	West North Central	Baccalaureate Colleges: arts and science focus	38
K	Mountain	D/PU: Doctoral/Professional Universities	51
L	South Atlantic	M3: Master's Colleges and Universities – smaller programs	18
M	New England	M1: Master's Colleges and Universities – larger programs	48
N	East North Central	D/PU: Doctoral/Professional Universities	42

Region in the U.S. is based on the U.S.A. Census division classification (Census, Regions, and Divisions of the United States, 2013).

construct validity was also established using Rasch analysis. Cronbach's alpha was used to assess reliability. A high alpha value indicates a high correlation among items and an acceptable Cronbach's alpha value is 0.70–1.00 (Nunnally & Bernstein, 1994). The reliability was also evaluated using person separation and item separation reliability (Bond & Fox, 2015).

Results and discussion

Inventory development

Identifying alternate conceptions

The responses of the participants to the open-ended questions were analyzed to identify common alternate conceptions. The most commonly occurring alternate conceptions were used as distractors when designing the items for the pilot version of the FC-ORMI.

Student interviews

The open-ended interviews with students helped to understand their thought processes while answering the questions and gave further information on alternate conceptions they possess. An example of an item where students were choosing the correct option for the answer tier but the incorrect option for the response tier is presented in Figure 2.

When the question was presented in an open-ended format, 14 out of 15 students who were given the question chose option (a) for the answer tier. This participant's (SS008) response for why 1 is the most stable indicates a fragmented understanding:

"It's because a positive charge is most stable around the most amount of offshoots. I don't know what that is called. The higher the degree is the more stable it is so the positive charge would want to move towards that because it's more stable." SS008

When presented with the above question the same 14 students chose option (a) for the answer tier but they were not sure about the response tier. These results suggest that students know that tertiary carbocations are the most stable but their reasoning is flawed. Of the 14 participants, seven chose the correct answer for the response tier but they were not confident about their choice. These results indicate that students have difficulty understanding fundamental concepts needed for understanding organic reaction mechanisms. Students seem to remember certain details due to either memorization or repeated exposure to the concept but they have difficulty with providing reasoning.

The alternate conceptions developed by students identified from administration of the open-ended questions and student interviews are given in Table 2. These identified alternate conceptions were used to develop items for the pilot version of the FC-ORMI.

Which is most stable?

- (a) 1
- (b) 2
- (c) 3

What is the reason for your answer to the above question?

- (a) Induction
- (b) Charge is centralized
- (c) It is the most symmetrical
- (d) Positive charge is at the end

Figure 2: Item assessing stability of intermediates used in student interviews.

Table 2: List of concepts and alternate conceptions identified.

Concepts	Alternate conceptions
Valence electrons	Valence electrons equal the number of protons Valence electrons are the number of unpaired electrons in the outermost shell
Lewis structures	Nitrogen atoms always form three bonds and carbon atoms form four bonds
Hybridization	Hybridization depends on the type of bond - single, double, triple Three electron groups indicate sp^3 hybridization
Molecular geometry	Double bond affects the geometry Lone pairs on carbon affect the geometry
Bonding	All bonds close to triple bonds are strong
Electron density/polarity	All electronegative atoms are electrophilic
Acids/bases	More electronegative leads to higher acidity Lewis base is a proton acceptor
Electrophiles/nucleophiles	Electrons flow from region of deficiency to region of density Sodium is positive and wants electrons so it is the electrophile The direction of electron flow
Resonance/induction effects	Best contributor to an overall resonance hybrid is the one that shows the movement of electrons and charges are spread out double bonds in the middle make more resonance structures Fluorine atoms contribute more to resonance
Stability of intermediates	When the charge is on a central atom it stabilizes the carbocation When the carbocation is symmetrical it is stable In a carbonyl group, electrons can go to the carbon and oxygen atoms want to donate electrons

Pilot and beta administration

The pilot version of the FC-ORMI contained 50 items. The difficulty and discrimination values for all items were computed (see Supplementary Material). The results indicated that only seven items (9, 19, 20, 21, 29, 44, and 46) had adequate difficulty and discrimination indices and could be retained without modification.

The remaining items were further analyzed by distractor analysis. For a distractor to be effective it should represent a plausible alternate conception; if a distractor does not attract any respondents then it may be implausible and not a good distractor (Bandalos, 2018). The data were analyzed in terms of the percentage each option was selected by the participants. An example of an item with a questionable distractor is shown in Figure 3.

The option (d) was under-utilized and hence was eliminated since it is not plausible and therefore not a good distractor. Similarly, distractor analysis was conducted for all the other items to help modify the items. The poorly performing items on the pilot version were modified or omitted based on results from the pilot administration to create the beta version of the FC-ORMI.

Which statement best describes the geometry of the indicated atom?

- (a) Trigonal planar (52.3%) (*)
- (b) Trigonal pyramidal (18.3%)
- (c) Bent (12.8%)
- (d) Linear (2.8%)
- (e) Tetrahedral (13.8%)

Figure 3: Item assessing molecular geometry with a questionable distractor.

The difficulty and discrimination indices for the 27 items on the beta version of the FC-ORMI were computed (see Supplementary Material). The results indicate that 16 out of the 27 items have adequate difficulty and discrimination values. The remaining items were analyzed using item difficulty and discrimination values. Distractor analysis was utilized to further analyze the items and the problematic items were modified or deleted based on item analysis to give the 26-item alpha version of the FC-ORMI.

Alpha administration

Psychometrics at the test level

The difficulty and discrimination indices for the 26 items on the alpha version of the FC-ORMI were computed (see Supplementary Material). The scatter plot depicting the difficulty and discrimination values for the 26 items is presented in Figure 4.

The results indicate that 20 of the 26 items have both difficulty and discrimination indices within the acceptable ranges. The six items that are out of the acceptable range for difficulty and discrimination are two-tier items where the linked pair is performing well. Therefore useful information regarding students' alternate conceptions can be obtained from these items.

Item 25 has a discrimination index of 0.23 and a difficulty index of 0.32 which suggests that the item is difficult and not discriminating well. Item 25 is the response tier to item 24; both items are shown in Figure 5.

Figure 4: Scatter plot showing difficulty (P) and discrimination (D) values for all 26 items on the alpha version of the FC-ORMI. Only the poorly-performing items have been labelled. Items 1, 4, 5, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, and 26 are within the rectangle which represents the accepted values of difficulty and discrimination.

24. Which of the following is the most stable?

- (a) 1 (54.4%) (*)
 (b) 2 (18.7%)
 (c) 3 (26.8%)

25. What is the reason for your answer to question 24?

- (a) The alkyl groups are electron donating. (27.0%) (*)
 (b) The carbocation is most symmetrical. (32.4%)
 (c) The carbocation can easily grab electrons. (40.6%)

Figure 5: Items on the alpha version of the FC-ORMI assessing the stability of intermediates.

Item 24 is functioning well and 54.4% of the participants chose the correct answer (a). Interestingly, 26.8% of the participants chose (c) which is the least stable carbocation. Item 25 is the response tier and based on the percentages for each response it can be seen that students are not able to effectively reason out their answers. Items 24 and 25 assess the concept of stability of intermediates. The results from the pilot and beta administrations indicated that this was a concept that students were not understanding thoroughly. These results further indicate that students have a fragmented understanding of carbocation stability and the factors affecting this stability. Even though their psychometric quality is below the acceptable range it can be argued that items 24 and 25 give important information.

Unidimensionality

One assumption made in the Rasch model is that the scale is unidimensional which means that it is measuring a single latent trait which is a hypothetical or unobserved characteristic (Bond & Fox, 2015). In this study, the latent trait can be defined as the conceptual knowledge of fundamental concepts for understanding organic reaction mechanisms. Unidimensionality within the Rasch model is analyzed by principal component analysis (PCA) of residuals where if the observed raw variance for measures is greater than 20%, the data are said to be unidimensional. The PCA of residuals for the data obtained using the alpha version of the FC-ORMI shows that 25.3% of the raw variance was explained by the Rasch model suggesting that the FC-ORMI is measuring a unidimensional construct. The model assumes that the test is unidimensional, meaning that each item probes only the measured variable and not something else. The unidimensionality requirements are, of course, only an approximation since a completely unidimensional test is almost impossible to construct but it can be realized well enough for practical purposes.

Item difficulty compared to person ability

The location of a person along the measured variable is described by a measure called person ability, which gives information about the intensity of the measured variable that the person possesses. Each item in the test should require different intensities of the variable in question. Items are described by another measure called item difficulty. A good test should contain items of various difficulties, more or less equally spaced along the whole range of abilities of the targeting sample. The Rasch model places item difficulty and person ability on the same logit scale using a vertical plot called a Wright map. The Wright map for the data obtained using the alpha version of the FC-ORMI is shown in Figure 6.

The Wright map shows that the range of logit measures is from a low of -3 to a high of $+3$. The item difficulties range from about -1.3 logits for the easiest item to about $+2.4$ logits for the most difficult item. This range of logit measures might indicate that the data are not skewed. However, on closer examination it can be seen that some students have ability lower than item 1 (10.11) which is the easiest item. This indicates that the inventory is marginally difficult and skewed with respect to person ability. Though the inventory is slightly difficult, none of the items were estimated with high imprecision indicating that the items are well targeted to the student population.

Item fit

The fit of the items to the Rasch model was evaluated using infit (weighted) and outfit (unweighted) statistics which help in identifying problematic items. The infit and outfit statistics are shown in Table 3.

All the items have mean square (MNSQ) infit and outfit statistics within the acceptable range of 0.70–1.30 for a low stakes multiple-choice test (Bond & Fox, 2015). Since the mean squares for all the items are acceptable, the ZSTD values can be ignored. These results suggest that the items fit well to the dichotomous Rasch model and there are no discrepancies between the observations and expectations for the items on the alpha version of the FC-ORMI.

Table 3: Fit statistics for the alpha version of the FC-ORMI data.

Item No.	Measure	SE	Infit <i>MNSQ</i>	Infit <i>ZSTD</i>	Outfit <i>MNSQ</i>	Outfit <i>ZSTD</i>
(8.9)	1.64	0.11	0.93	-1.2	0.96	-0.3
(24.25)	1.46	0.10	1.01	0.2	1.18	1.8
(21.22)	0.80	0.09	0.98	-0.5	1.01	0.1
(6.7)	0.71	0.09	1.03	0.9	1.00	0.0
(18.19)	0.56	0.09	1.06	1.6	1.09	1.5
(12.13)	0.30	0.08	1.13	3.7	1.15	3.0
(16.17)	-0.11	0.08	0.93	-2.3	0.92	-1.9
(14.15)	-0.24	0.08	0.99	-0.2	1.02	0.4
20	-0.31	0.08	1.06	2.2	1.05	1.1
23	-0.48	0.08	0.87	-0.48	0.82	-4.2
26	-0.50	0.08	0.99	-0.2	1.02	0.3
5	-0.68	0.08	1.00	0.1	0.98	-0.4
(3.4)	-0.86	0.08	0.92	-2.7	0.86	-2.6
(1.2)	-0.94	0.08	1.01	0.5	1.03	0.5
(10.11)	-1.34	0.09	1.03	0.8	1.20	2.7
<i>M</i>	0.00	0.09	1.00	-0.1	1.02	0.1
<i>SD</i>	0.85	0.01	0.06	2.0	0.10	1.8

SE, standard error; *MNSQ*, mean square value; *ZSTD*, standardized z-score

The content validity was established by sending the instrument to organic chemistry faculty ($N = 8$) to enable them to revise the content if necessary for clarity, correctness, and relevance. One of the items on the inventory assessing the concept of electrophiles and nucleophiles did not have a response tier. The item is presented in Figure 7.

The feedback from the experts indicated that labeling the molecules by their names would not be appropriate since that would also test nomenclature, and the use of a simpler nucleophile in place of Na^+CN^- was recommended. Additionally, it was suggested that if the forward reaction was being considered then only the reactants need to be labeled and not the products. This would also help in setting up a response tier to get further information regarding students' understanding. The item was modified based on these suggestions as shown in Figure 8.

Which is the electrophile and which is the nucleophile in the following reaction?

Figure 7: Item assessing electrophiles and nucleophiles without a response tier.

Which is the electrophile in the following reaction?

What is the reason for your answer to the above question?

- (a) **1** is the electrophile because it contains the most electropositive atom.
 (b) **1** is the electrophile because it has the electrons which are needed to become an ion.
 (c) **2** is the electrophile because it contains the most electronegative atom.
 (d) **2** is the electrophile because it is the strongest base.

Figure 8: Modified item assessing electrophiles and nucleophiles.

The performance of students on the FC-ORMI items were compared to the performance of students on similar items on the ACS standardized general chemistry and organic chemistry exams. Of the 26 FC-ORMI items on the alpha version, 15 items had similarities with items on the ACS General Chemistry Exams (Form 2015, Form 2017) and ACS Organic Chemistry Exams (Form 2014, Form 2016). Most of the items on the FC-ORMI had comparable difficulty and discrimination indices to similar items on the ACS standardized exams. The FC-ORMI items were also compared to the anchoring concepts content map for undergraduate organic chemistry proposed by the ACS Exams Institute (Raker, Holme, & Murphy, 2013). It was found that all nine concepts under which the items were grouped are present in the anchoring concepts content map.

The PCA of residuals indicated that 25.3% of the raw variance was explained by the Rasch model suggesting the unidimensionality of the instrument. The PCA of residuals together with the item fit to the dichotomous Rasch model, suggests that the alpha version of the FC-ORMI is measuring the intended construct of students' conception of fundamental concepts for understanding organic reaction mechanisms. These results further establish the construct validity of the data obtained using the FC-ORMI.

Assessment of reliability

Several aspects of reliability were assessed. The internal consistency reliability was measured using Cronbach's alpha for the pilot version and beta version and the values obtained were 0.682 and 0.632, respectively, which are slightly lower than the accepted value of 0.7 although Cronbach's alpha is the lower bound of reliability compared to other forms of reliability (Nunnally & Bernstein, 1994). The Cronbach's alpha values were also obtained if items were deleted and this information was used together with results from the pilot and beta versions to eliminate items during the development of the inventory. The Cronbach's alpha value for the 26-item alpha version of the FC-ORMI was 0.712 which indicates that the items are correlated.

The item separation reliability was found to be 0.99 which is higher than the accepted value of 0.8, and hence it is adequate. The person separation reliability was found to be only 0.60 which is lower than the accepted value of 0.8. It is difficult to obtain reliability using Rasch analysis within the acceptable range and it is easier to obtain item separation reliability within the acceptable range rather than person separation reliability (Bond & Fox, 2015). The person separation reliability could be improved by increasing the number of items and creating more categories.

Limitations

The pilot version of the FC-ORMI contained 50 items. The large number of items could have resulted in a higher cognitive load for the participants. Though most of the items in the alpha version of the FC-ORMI were functioning well, six items did not have adequate difficulty and discrimination. However, important information regarding students' alternate conceptions can still be obtained from these items.

Conclusions

The purpose of this study was to develop and psychometrically assess the Fundamental Concepts for Organic Reaction Mechanisms Inventory (FC-ORMI). The validity and reliability of the data obtained using the FC-ORMI meet the acceptable standards suggesting that the instrument is functioning well to determine students' alternate conceptions of fundamental concepts for understanding organic reaction mechanisms.

Implications for teaching

The FC-ORMI could be used as a summative assessment by organic chemistry instructors to detect students' alternate conceptions on these fundamental concepts before they begin the study of organic reaction mechanisms and accordingly modify their course content to help students overcome difficulties with these concepts. Instructions for administration of the FC-ORMI are included in the Supplementary Material. Alternatively, specific items from the inventory could be used by instructors for formative assessments or as clicker questions. The information obtained using the FC-ORMI could also be useful to general chemistry instructors giving them information regarding the utility of general chemistry concepts in organic chemistry.

Additionally, the FC-ORMI could be useful for establishing the efficacy of instructional materials. Information regarding the type of textbook used for first-semester organic chemistry by the 14 different schools who administered the alpha version of the FC-ORMI were collected. The type of textbook used by individual schools are presented in Table 4 along with the average scores for each school.

The overall raw average on the alpha version of the FC-ORMI across 14 different schools was 55.5% with the highest being 73.8% and the lowest being 40.8%. Four out of the 14 schools (F, I, J, and N) use a textbook that is organized using a mechanistic approach. Schools F, I, and J had the three highest average scores with school I obtaining the highest average of 73.8%. An organic chemistry textbook where the content is organized by reaction mechanisms (Karty, 2018) is different from the traditional textbooks (Bruice, 2014; Carey, Giuliano, Allison, & Tuttle, 2020; Clayden, Greeves, & Warren, 2012; Klein, 2017; McMurry, 2018; Solomons, Fryhle, & Snyder, 2016; Wade & Simek, 2017) where the content is organized by functional groups. A number of different factors could be responsible for this difference in performance but one of the factors could be the textbook being used and this requires further investigation.

Future work

The development of a concept inventory is an iterative process and even the alpha version of the FC-ORMI can be improved in the future. The Rasch analysis indicated that the inventory is marginally difficult for students and this could be addressed by modifying the items. More items for each concept could be added to increase the person separation reliability. Additionally, Differential Item Functioning (DIF) could be used to compare the performance on the FC-ORMI by gender or textbook used to gain further information on the factors contributing to student understanding of fundamental concepts.

Table 4: Type of textbook used by schools in alpha administration.

School	Type of textbook used	Average scores
A	Functional groups	48.3%
B	Functional groups	49.7%
C	Functional groups	46.6%
D	Functional groups	56.8%
E	Functional groups	53.9%
F	Reaction mechanisms	66.8%
G	Functional groups	51.8%
H	Functional groups	47.9%
I	Reaction mechanisms	73.8%
J	Reaction mechanisms	66.8%
K	Functional groups	60.3%
L	Functional groups	40.8%
M	Functional groups	57.5%
N	Reaction mechanisms	57.8%

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

- Anderson, T. L., & Bodner, G. M. (2008). What can we do about “Parker”? A case study of a good student who didn’t “get” organic chemistry. *Chemistry Education Research and Practice*, 9(2), 93–101.
- 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.
- Atkinson, M. B., Popova, M., Croisant, M., Reed, D. J., & Bretz, S. L. (2020). Development of the reaction coordinate diagram inventory: measuring student thinking and confidence. *Journal of Chemical Education*, 97(7), 1841–1851.
- Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. New York, NY: The Guilford Press.
- Barbera, J. (2013). A psychometric analysis of the chemical concepts inventory. *Journal of Chemical Education*, 90(5), 546–553.
- Bhattacharyya, G. (2013). From source to sink: mechanistic reasoning using the electron-pushing formalism. *Journal of Chemical Education*, 90(10), 1282–1289.
- Bhattacharyya, G. (2014). Trials and tribulations: student approaches and difficulties with proposing mechanisms using the electron-pushing formalism. *Chemistry Education Research and Practice*, 15(4), 594–609.
- Bhattacharyya, G., & Bodner, G. M. (2005). “It gets me to the product”: How students propose organic mechanisms. *Journal of Chemical Education*, 82(9), 1402.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. New York, NY: Routledge.
- Boone, W. J., & Scantlebury, K. (2006). The role of rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, 90(2), 253–269.
- Boone, W. J., Townsend, J. S., & Staver, J. (2011). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: an exemplar utilizing STEBI self-efficacy data. *Science Education*, 95(2), 258–280.
- Brandriet, A. R., & Bretz, S. L. (2014). The development of the redox concept inventory as a measure of students’ symbolic and particulate redox understandings and confidence. *Journal of Chemical Education*, 91(8), 1132–1144.
- Bretz, S. L., & Linenberger, K. (2012). Development of the enzyme-substrate interactions concept inventory. *Biochemistry and Molecular Biology Education*, 40(4), 229–233.
- Bretz, S. L., & Murata Mayo, A. V. (2018). Development of the flame test concept inventory: measuring student thinking about atomic emission. *Journal of Chemical Education*, 95(1), 17–27.
- Brown, C. E., Hyslop, R. M., & Barbera, J. (2015). Development and analysis of an instrument to assess student understanding of GOB chemistry knowledge relevant to clinical nursing practice. *Biochemistry and Molecular Biology Education*, 43, 13–19.
- Bruice, P. Y. (2014). *Organic chemistry*. Boston: Pearson.
- Carey, F. A., Giuliano, R. M., Allison, N. T., & Tuttle, S. L. B. (2020). *Organic chemistry*. New York, NY: McGraw-Hill Education.
- Census, Regions, and Divisions of the United States (2013). Retrieved from https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf.
- Chu, H. E., Treagust, D. F., Yeo, S., & Zadnik, M. (2012). Evaluation of students’ understanding of thermal concepts in everyday contexts. *International Journal of Science Education*, 34(10), 1509–1534.
- Clayden, J., Greeves, N., & Warren, S. G. (2012). *Organic chemistry*. Oxford: Oxford University Press.
- Dick-Perez, M., Luxford, C. J., Windus, T. L., & Holme, T. (2016). A quantum chemistry concept inventory for physical chemistry classes. *Journal of Chemical Education*, 93(4), 605–612.
- Doran, R. (1980). *Basic measurement and evaluation of science instruction*. Washington, DC: National Science Teacher Association.
- Ebel, R. L. & Frisbie, D. A. (2006). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Ferguson, R., & Bodner, G. M. (2008). Making sense of the arrow-pushing formalism among chemistry majors enrolled in organic chemistry. *Chemistry Education Research and Practice*, 9(2), 102–113.
- Grove, N. P., Cooper, M. M., & Cox, E. L. (2012). Does mechanistic thinking improve student success in organic chemistry? *Journal of Chemical Education*, 89(7), 850–853.
- Grove, N. P., Cooper, M. M., & Rush, K. M. (2012). Decorating with arrows: toward the development of representational competence in organic chemistry. *Journal of Chemical Education*, 89(7), 844–849.
- Hadenfeldt, J. C., Bernholt, S., Liu, X., Neumann, K., & Parchmann, I. (2013). Using ordered multiple-choice items to assess students’ understanding of the structure and composition of matter. *Journal of Chemical Education*, 90(12), 1602–1608.

- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38–47.
- Kaplan, R. M., & Saccuzzo, D. P. (1997). *Psychological Testing: Principles, Applications, and Issues*. Pacific Grove, CA, Brooks/Cole.
- Karty, J. (2018). *Organic chemistry: Principles and mechanisms*. New York: W. W. Norton.
- Kelley, T. L., (1939). The selection of upper and lower groups for the validation of items. *Journal of Educational Psychology*, 30, 17–24.
- Klein, D. R. (2017). *Organic chemistry*. Hoboken: John Wiley & Sons, Inc.
- Kraft, A., Strickland, A. M., & Bhattacharyya, G. (2010). Reasonable reasoning: multi-variate problem-solving in organic chemistry. *Chemistry Education Research and Practice*, 11(4), 281–292.
- Livescribe. (2016). Retrieved from <https://www.Livescribe.com/>.
- Luxford, C. J., & Bretz, S. L. (2014). Development of the bonding representations inventory to identify student misconceptions about covalent and ionic bonding representations. *Journal of Chemical Education*, 91, 312–320.
- McClary, L. M., & Bretz, S. L. (2012). Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *International Journal of Science Education*, 34(15), 2317–2341.
- McMurry, J. (2018). *Organic chemistry*. Boston: Cengage Learning.
- Mulford, D. R., & Robinson, W. R. (2002). An inventory for alternate conceptions among first-semester general chemistry students. *Journal of Chemical Education*, 79(6), 739.
- Nedungadi, S., & Brown, C. E. (2020). Thinking like an electron: concepts pertinent to developing proficiency in organic reaction mechanisms, *Chemistry Teacher International*, 3(1), 9–17.
- Nedungadi, S., Paek, S. & Brown, C. (2019). Utilizing Rasch analysis to establish the psychometric properties of a concept inventory on concepts important for developing proficiency in organic reaction mechanisms. *Chemistry Teacher International*. <https://doi.org/10.1515/cti-2019-0004> [Epub ahead of print].
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Othman, J., Treagust, D. F., & Chandrasegaran, A. L. (2008). An investigation into the relationship between students' conceptions of the particulate nature of matter and their understanding of chemical bonding. *International Journal of Science Education*, 30(11), 1531–1550.
- Pentecost, T. C., & Barbera, J. (2013). Measuring learning gains in chemical education: a comparison of two methods. *Journal of Chemical Education*, 90(7), 839–845.
- Raker, J., Holme, T., & Murphy, K. (2013). The ACS exams institute undergraduate chemistry anchoring concepts content map II: organic chemistry. *Journal of Chemical Education*, 90, 1443–1445.
- Richardson, J. (2004). Concept inventories: Tools for uncovering STEM students' misconceptions. *Assessment and Education Research* [Online].
- Solomons, T. W. G., Fryhle, C. B., & Snyder, S. A. (2016). *Organic chemistry*. Hoboken, NJ: Wiley.
- Taskin, V., Bernholt, S., & Parchmann, I. (2015). An inventory for measuring student teachers' knowledge of chemical representations: design, validation, and psychometric analysis. *Chemistry Education Research and Practice*, 16(3), 460–477.
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159–169.
- Villafañe, S. M., Bailey, C. P., Loertscher, J., Minderhout, V., & Lewis, J. E. (2011). Development and analysis of an instrument to assess student understanding of foundational concepts before biochemistry coursework. *Biochemistry and Molecular Biology Education*, 39(2), 102–109.
- Wade, L. G., & Simek, J. W. (2017). *Organic chemistry*. Harlow: Pearson.
- Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using Rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education*, 89(3), 335–345.
- Wren, D., & Barbera, J. (2013). Gathering evidence for validity during the design, development, and qualitative evaluation of thermochemistry concept inventory items. *Journal of Chemical Education*, 90(12), 1590–1601.
- Wren, D., & Barbera, J. (2014). Psychometric analysis of the thermochemistry concept inventory. *Chemistry Education Research and Practice*, 15(3), 380–390.