Deeper look at question categories, concepts, and context covered: Modified module analysis of quantum mechanics concept assessment

James Wells¹,¹ Homeyra Sadaghiani²,² Benjamin P. Schermerhorn,^{3,2}
Steven Pollock⁴,⁴ and Gina Passante³

¹*Department of Physics, University of Connecticut, Storrs, Connecticut 06269, USA*

²*Department of Physics & Astronomy, California Polytechnic University Pomona,
Pomona, California 91768, USA*

³*Department of Physics, California State University, Fullerton, Fullerton, California 92831, USA*

⁴*Department of Physics, University of Colorado, Boulder, Boulder, Colorado 80309, USA*

(Received 13 April 2021; accepted 29 June 2021; published 30 August 2021)

Concept inventories are commonly used tools in physics education research (PER) for evaluating teaching methods. Student responses to concept inventories have been studied using classical test theoretical methods such as factor analysis and network methods such as module analysis. The results of these studies have been used to evaluate the instruments and to find aggregate patterns that may be used to infer how students' think about the concepts on the inventory. In this work, we apply modified model analysis using partial correlations, a network based method using partial correlations, to evaluate the Quantum Mechanics Concept Assessment (QMCA). The QMCA is designed to measure student understanding of upper-division quantum mechanics in both the spatial wave function and the spins contexts. This represents the first time a network-based analysis has been applied to an upper-division concept inventory. The modules we found are related to the structure of the instrument. There were two broad classes of modules: those that connected responses (either both correct or both incorrect) to isomorphic pairs questions written to probe the same concept in the spatial wave function and the spin contexts. The second class of modules were explain pairs, where the first question asked about a concept and the second question asked students to explain their choice, again either both correct or both incorrect. This structure resembles the structure of the Force Concept Inventory where there are a common correct and a common incorrect pair for many concepts, compared to the Force and Motion Conceptual Evaluation, which has larger and more varied communities reflecting a smaller number of topics.

DOI: [10.1103/PhysRevPhysEducRes.17.020113](https://doi.org/10.1103/PhysRevPhysEducRes.17.020113)

I. INTRODUCTION

Well-designed, valid, and reliable assessment tools can help inform instructors by casting light on student ideas about key concepts. Conceptual assessments developed based on physics education research (PER) are widely used to evaluate transformed pedagogies, to develop more effective curriculum, and more generally, to guide efforts towards the systematic improvement of physics instruction on a range of topics in lower- and upper-division courses. Compared to introductory physics topics, inventories on upper-division topics are relatively limited. There are, however, a number of quantum mechanics concept inventories focused on a variety of topics [1–5]. Most diagnostic

tools are developed based on expert classroom observations and research findings on student common ideas or challenges. However, it is not the case that scores on a concept inventory directly correspond to how well students understand those common ideas. If the instrument includes blocked problems—such as items that are based on the same description of a situation, known as common stem items—the items are no longer independent and scoring them separately does not reflect that. Further, patterns in student responses show that items that experts understand as testing equivalent concepts are not always interpreted that way by students [6]. We had previously developed a 31-item multiple choice instrument [the Quantum Mechanics Concept Assessment (QMCA) [7]] to gauge student learning in upper-division quantum mechanics courses. By comparing student learning across different populations and institutions, the QMCA provides instructors and researchers with a standard measure of the effectiveness of different curricula or teaching strategies on student conceptual understanding

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

of quantum mechanics. QMCA questions are motivated by previous research on teaching and learning quantum mechanics. As much as possible, distractors replicated common student reasoning that repeatedly appeared in our interviews and existing literature [8–11]. The QMCA is given postinstruction and contains questions on five main (nonorthogonal) topics of quantum mechanics: measurement, the time-independent Schrödinger equation, wave functions and boundary conditions, time evolution, and probability density. The original QMCA questions [V 5.7] addressed the traditional upper-division context of the spatial wave function of a single particle in a one-dimensional infinite potential well. Since then, a growing number of physics instructors have adopted a spins first approach [7] to teaching undergraduate quantum courses. For the last four years, we have focused on research and curriculum development tailored to a spins first approach [12]. We have developed a wide range of curricular material including prelecture assignments, tutorials, concept tests, lecture notes, homework, and exam questions for this teaching approach. These developments have motivated us to add nine questions in the context of spin half systems to the original QMCA to make it a more versatile tool for various teaching paradigms. The newly added spins items were intended by the authors of the QMCA to be isomorphic to existing spatial wave function items, requiring the same formal reasoning in the new context. We will refer to these as isomorphic question pairs throughout the paper, however this does not necessarily imply that students answer these questions similarly. In the development process of the QMCA, test validity (the degree to which the instrument measures what it is intended to measure) was established by multiple rounds of experts' and students' feedback and interviews. The QMCA reliability (the extent to which the instrument gives results that are consistent) was verified by various statistical analysis including item difficulty index, item discrimination, Kuder-Richardson reliability index, and Ferguson's delta. In this paper we analyze the latest version of the QMCA (Version 6.2.2), containing the spins context items, [7] using modified model analysis using partial correlations (MMA-P) [13]. Modified module analysis is a network analytic technique for extracting groups of answers that are commonly chosen together, known as modules [14,15]. MMA-P uses partial correlations to measure the correlations between responses with the effect of overall score removed.

This study aims to answer the following questions:

RQ1: Are student responses to the spins context items correlated to their responses to the items in the position context that experts intended to be isomorphic?

RQ2: What do the patterns in student responses detected by MMA reveal about the structure of the QMCA?

In addition to information about the instrument, student responses may also suggest future avenues of research into

student beliefs about quantum mechanics or question design in different quantum mechanical contexts.

II. QMCA STRUCTURE AND DEVELOPMENT

During the development of the QMCA, we (H.S. and S.P.) chose to include questions about five topics based on faculty-consensus learning goals and on the typical structure of introductory quantum courses. The five question categories were intended to broadly represent the essential topics in undergraduate junior- or senior-level quantum mechanics courses, based on learning objectives discussed by faculty experts. Each dimension is probed by more than one question on the same subject to enhance the test's validity and reliability [7].

The QMCA categories were labeled measurement (M), time-independent Schrödinger equation (TISE), time evolution (TE), wave functions (W), and probability (P). In brief, M questions include interpretation of coefficients as probabilities when expanding in a basis and collapse of states after measurement. TISE questions involve forms and qualitative features of solutions for wave functions. TE includes use of time-dependent phases when expanding the state in the energy basis, and questions about the time evaluation of position and energy measurements as well as expectation values. The W category includes questions about the complex-valued nature of states, and connections to spatial probability densities. The P category includes a variety of questions about the probabilistic interpretation of wave functions and expansion coefficients (both time dependent and independent). There is considerable overlap between these categories, with multiple questions having labels in more than one category (more details can be found in Ref. [7]).

A. Addition of spin questions and a new categorization

The original QMCA questions were explicitly framed in the context of a particle in a 1D infinite well. Because of the changing nature of quantum mechanics instruction in the last five years, there has been a need to adjust the QMCA to match the evolution of instruction. To this end, we formulated variants to nine existing questions to create paired, isomorphic questions explicitly in the context of spin- $\frac{1}{2}$ particles, expressed in Dirac notation. We developed the new variants and their distractors in an iterative process involving post-test interviews and examination of students' reasoning on closely matched exam questions. In the process, some of the older position-context questions were slightly reworded to make phrasing and options as close as possible between the spin and position contexts. Our preliminary comparative study [16] of over 100 students at 3 institutions, all taught in spins-first curricula showed strong consistency in student responses across question context. This work will dig deeper into the correlations between the spins and position contexts.

B. Structure within the QMCA

Physics concept inventories often include blocked problems [6,14]. Blocked problems are problems that are connected either directly or through a common problem stem.

The QMCA contains two types of blocked problems. The first type are the questions designed to be isomorphic between the wave function and spins contexts. There are nine such isomorphic pairs on the revised QMCA.

The second type of blocked problems are explain-pair items. In these six pairs, the second question asks students to select the best explanation for their answer to the first question. In four of the pairs, the first question is true-false. Two of the explain pairs are made up of items that are also part of wave function–spin isomorphic pairs: $Q4 - Q5$ and $Q25 - Q26$, where $Q4 - Q25$ and $Q5 - Q26$ are isomorphic pairs.

III. RESEARCH STUDY

A. Study sample

Data for this study were collected from upper-division quantum mechanics courses at five different institutions. All quantum mechanics courses that contributed results used a “spins-first” instructional paradigm following McIntyre’s *Quantum Mechanics* textbook [17], and at our institutions, our new spins-first curriculum and interactive methods were implemented [12].

The QMCA was administered at the end of the first semester of quantum mechanics at each institution. At one location, a paper version of the QMCA was administered during standard class time. The other schools used an online format where students answered the assessment outside of class time. A recent study, which included QMCA data along with other upper-division assessments, found no significant change in population average performance for online assessment that would threaten our ability to interpret performance, or compare it to paper-based implementations [18].

The data from all five institutions were aggregated into a single sample of 281 students. The total scores on the 38-question exam ranged from 16% to 97%, with a mean of 59% (median 61%); see Fig. 1. The sample size of each individual institution is too small to be confident in the disaggregated results, so we only report the aggregate results.

B. Modified module analysis using partial correlations

Modified module analysis using partial correlations is a network analytic technique for extracting groups of answers that are commonly chosen together, known as modules, from concept inventories [13–15,19,20]. A popular method of studying concept inventories is to use factor analysis [21–24]. Factor analysis is applied at the item level and finds correlations between correctly

FIG. 1. The empirical cumulative distribution function displays the percent correct on the QMCA on the x axis and the percent of the sample with that score or less on the y axis. The data have a slight right skew, but are approximately normal.

answered items. The incorrect responses, which have been chosen because they represent different common student ideas, are all treated as a single category: incorrect. In contrast, MMA-P correlates student responses, so each distractor is considered separately. Thus, it can result in a more fine-grained understanding of student ideas.

The basic method of MMA-P is to create a network of responses to the concept inventory, then use a community detection algorithm (CDA) to find modules. A network is a representation of a dataset as nodes connected to one another by edges. In the case of MMA-P, each node represents an answer choice, and the edges connect responses commonly chosen together. Edges can also be valued or weighted, to represent the strength or intensity of the connection between the nodes. In MMA-P, the edges are valued according to the partial correlation between the items, controlling for the total scores of the test takers. Edges representing correlations that are not significant at the Bonferroni-corrected $\alpha < .05$ level, negative correlations (typically among responses to the same item), and positive correlations below a given threshold (in this case $r < 0.1$, the small effect size threshold) are discarded. The effect size of a correlation is the correlation value itself; the effect size thresholds are 0.1 for a small effect, 0.3 for a medium effect, and 0.5 for a large effect. All data preparation and analysis were done in the R language for statistical programming, version 3.5.2 [25] and the RStudio development environment [26].

Once the network is formed, a CDA is used to determine which nodes cluster together to form coherent subnetworks or communities within the larger network. There are many CDAs that have been developed to identify subnetworks [27]. MMA-P uses the fast-greedy algorithm [28] which is a CDA designed to maximize the modularity of the

network [29]. Modularity is a measure of the difference between the identified set of clusters and a set of clusters chosen at random on the same network. The modularity is expressed as a modularity score between zero and one that indicates the community structure in the network. A modularity score of zero means the network has no substructure, i.e., no division of the network into subgroups is better than random assignment. A modularity score of one indicates a highly modular structure. A division that results in a high modularity score has many more edges within the subgroups than between the subgroups, compared to a baseline measurement for nodes randomly assigned to groups in the particular network. In general, a high value of modularity is interpreted to mean that the network structure was created by a process other than chance. In the context of MMA-P, we assume that the process creating the structure is the students selecting the responses that best match their understanding of the physical concepts.

There are several potential difficulties with applying the process of network creation and CDA only to the observed network. First, the division of the network into clusters by the fast greedy CDA is not deterministic and could give slightly different results on different applications. Because the number of possible communities may be too large to feasibly compute the best division among all possible divisions, fast greedy does a series of local optimizations. For example, the selection of a division of existing communities if two or more divisions improve the modularity score by the same amount may result in a different global value of modularity.

Second, there is no information about the population that the sample of students was chosen from which is needed to generalize beyond this group of students. Thus, we cannot use parametric statistics such as the standard deviation, because it is unknown if the underlying distribution is normal. To understand whether the result is plausible, it is necessary to have a distribution of groupings to compare with. Therefore, one-thousand networks are created from bootstrapped samples of students and then the CDA is applied to each network.

Bootstrapping is a statistical technique used when the underlying distribution of the population is unknown. Instead of simply measuring the correlations between responses in the collected sample, they are measured on many samples that are the same size as the collected sample, generated by random selection with replacement from the original sample. With bootstrapping, we can compare the observed network to the population of randomly selected samples from the data, created using observed response frequencies. The bootstrapping technique suppresses uncommon item pairings in favor of connections of more commonly chosen items.

A network is formed from the partial correlation matrix of the bootstrapped and communities were found using the

implementation of the fast-greedy algorithm in the igraph package cite [30]. Only responses that are assigned the same community in 800 of the 1000 samples are grouped together in the module network. A final network is formed with responses as the nodes connected by edges whose weight is equal to the number of times the two responses were in the same community in the thousand bootstrapped samples.

The QMCA has questions with two, three, four, and five possible responses. Any response that was selected by ten or fewer respondents was removed, to reduce the noise from correlations between infrequently chosen responses. One-thousand samples of 281 students were bootstrapped using the dplyr package [31] and the correlations and their p-values were measured using the psych package [32]. Only correlations $r > 0.1$ (a small effect size) that were significant to the $\alpha = 0.05$ level with a Bonferroni correction are included in the network created from each sample.

IV. RESULTS

In this section we analyze student responses to the QMCA using two different techniques. First, addressing research question 1, we examine only those questions which were intended by the authors to be isomorphic across spin and wave function contexts. Second, addressing research question 2, we use MMA-P to analyze the full instrument in an effort to explore the structure of the QMCA.

A. Isomorphic spin and wave function questions

One goal of the QMCA is to study student understanding of quantum mechanics in both the spin and wave function contexts. As a result, there are nine pairs of questions that intend to ask fundamentally the same question but in the wave function and spin contexts, an example of such a pairing is shown in Fig. 2.

1. Partial correlations between paired correct responses

Table I shows the partial correlation between the correct responses for isomorphic questions, controlling for overall score, as well as a brief description of the questions. Students tend to answer items in five of the isomorphic pairs either both correctly or both incorrectly. The correct response in these five isomorphic pairs are all correlated at 0.29 or greater. Four of these pairs of questions ask specifically about measurements where the state is written in the basis of the operator being measured. The medium effect size (0.30) of the correlation between all but one of these responses (see Table I) supports the claim that these items are functioning as intended by the authors of the QMCA. For example, a student's ability to predict the maximum value that can be measured from a superposition state in one context (see Fig. 2) is predicted by their ability to do so in the other context. The four pairs that have

For questions 1-5, consider a particle in a one-dimensional infinite square well with width a , centered at $a/2$. There normalized energy eigenstate wave functions are $\psi_n(x)$ with energies E_n . The particle starts in a state given by

$$\Psi(x, t=0) = \sqrt{\frac{4}{5}}\psi_1(x) + \sqrt{\frac{1}{5}}\psi_2(x)$$

1. What is **the maximum value** that could be measured for the energy of the system?

- A. $\frac{4}{5}$
- B. E_1
- C. E_2
- D. $\frac{4}{5}E_1 + \frac{1}{5}E_2$
- E. Any E_n could be measured.

For this page, consider a spin-1/2 particle. The normalized eigenstates of S_z (the z -component of spin) are labeled as $|+\rangle$ and $|-\rangle$, with corresponding eigenvalues $\pm\hbar/2$. Consider a particle which starts in the quantum state given by

$$|\psi\rangle = \sqrt{\frac{1}{5}}|+\rangle + \sqrt{\frac{4}{5}}|-\rangle$$

22. What is **the maximum value** that could be measured for the energy of the system?

- A. $\frac{4}{5}$
- B. $-\hbar/2$
- C. $+\hbar/2$
- D. $\frac{1}{5}\hbar/2 + \frac{4}{5}(-\hbar/2)$
- E. $\sqrt{\frac{1}{5}}\hbar/2 + \sqrt{\frac{4}{5}}(-\hbar/2)$

FIG. 2. Example of isomorphic questions. In the spatial wave function context students work with the energy eigenstates in the infinite square well. In the spins context, students work with spin-1/2 and spin-1 systems. The states are both written with identical coefficients and the questions ask for the same conceptual information.

correlations above 0.30 are identified as modules by the MMA-P analysis, as discussed in Sec. IV B 1.

The correct responses in the other four isomorphic pairs are uncorrelated once the total score is taken into account (see Table I). Three of the isomorphic pairs with partial correlations below 0.30 are related to repeated measurements of energy and position or of S_z and S_x , respectively. Further, items 4 and 5 and items 25 and 26 are explain pairs which are isomorphic to each other, i.e., items 4 and 25 and items 5 and 26 form isomorphic pairs. The correct responses to these questions have a nonsignificant correlation

independent of the total score with a small or lower effect size, which may indicate that students do not recognize these items as isomorphic or that they have difficulty transferring the concepts between contexts.

2. Partial correlations between paired incorrect responses

In addition to the correlations between correct answers, we found strong correlations between several distractors in the isomorphic questions that represent the same incorrect ideas, controlling for overall score. One possible interpretation is that because these ideas are held in both

TABLE I. Partial correlation values between correct responses on paired questions controlling for the total score. Labels refers to the module identification in Fig. 3. If no label is provided the pairing does not appear as a module in the MMA-P analysis Fig. 3. Question numbers refer to QMCA version 6.2.2 with the letter indicating the answer option.

Question summary	Wave function (% correct)	Spin (% correct)	Phi correlation (p value)	Label (item and response)
Does the addition of a relative phase affect the probability of measuring energy (or spin)?	77%	76%	0.52(<0.001)	Iso4 (9A, 30A)
What is the normalized state after the most probable value is measured?	83%	85%	0.42(<0.001)	Iso1 (2C, 23C)
Does $\hat{H}\psi(x)$ (or $\hat{Q} \psi\rangle$) describe the state after a measurement of $\hat{H}$ (or $\hat{Q}$)? (T/F Question)	50%	46%	0.33(<0.001)	Iso3 (7B, 28B)
What is the state after a measurement of $\hat{H}$ (or $\hat{S}_z$)?	60%	72%	0.32(<0.001)	Iso2 (6D, 27D)
What is the maximum value that can be measured from a superposition state?	79%	88%	0.29 (0.01)	(1C, 22C)
Is there a definite energy (or spin) value for a superposition state? (T/F Question)	92%	85%	0.16 (1.00)	(8B, 29B)
Choose an explanation for your answer to whether the probabilities of sequential measurements of position then energy (or of $\hat{S}_x$ then $\hat{S}_z$) outcomes depend on time?	48%	41%	0.12 (1.00)	(5C, 26C)
Do the probabilities of sequential measurements of position then energy (or of $\hat{S}_x$ then $\hat{S}_z$) outcomes depend on time? (Y/N Question)	64%	59%	0.06 (1.00)	(4B, 25B)
What are the possible values of repeating measurements for noncommuting operators? (Energy and position or $\hat{S}_x$ and $\hat{S}_z$)	23%	55%	0.05 (1.00)	(3D, 24D)

TABLE II. Statistically significant partial correlations at the $p < 0.05$ level between commonly selected ($>10\%$) incorrect responses on paired questions. Labels refers to the module identification in Fig. 3. If no label is provided the pairing does not appear as a module in the MMA-P analysis Fig. 3. Question numbers refer to QMCA version 6.2.2 with the letter indicating the answer option.

Authors' view of the conceptual error in these isomorphic distractors	Wave function (% chosen)	Spin (% chosen)	Phi correlation (p value)	Label (item and response)
The addition of a relative phase to a state written in the energy (S_z) basis affects the probability of measuring energy (spin)	16.5%	13.8%	0.63(<0.001)	Iso8 (9C, 30C)
Treating a measurement outcome as equivalent to acting with corresponding operator	6.8%	5.3%	0.45(<0.001)	Iso5 (1D, 22D)
Making a measurement has no effect on the state	7.5%	3.9%	0.36(<0.001)	Iso6 (2E, 23E)
The state after measuring operator $\hat{A}$ on state $\psi(x)$ ($ \psi\rangle$) is given by $\hat{A}\psi(x)$ ($\hat{A} \psi\rangle$) (T/F question)	47.8%	32.2%	0.34(<0.001)	Iso7 (7A, 28A)
Confusing probability with eigenvalue	2.5%	1.8%	0.31 (0.001)	(1A, 22A)
Failing to re-normalize the state after a measurement	5.7%	6.8%	0.30 (0.003)	(2A, 23A)
The state after measurement is undetermined but will be a product of eigenvalue and the eigenstate (instead of just the eigenstate)	27.2%	18.3%	0.29 (0.001)	(6C, 27C)

contexts, these may be more strongly held incorrect ideas than other responses. Notably, the isomorphic items with the largest correlations between the correct responses also had strong correlations between a pair of distractors (see Table II). This pattern of strong correlations between distractors in two contexts that experts perceive to be equivalent is consistent with the predictions of the misconceptions paradigm. Misconceptions are durable, context-independent alternative ideas about physics topics [33]. However, MMA-P cannot determine the theory of knowledge which best represents how students approach these items. To answer that question, qualitative study beyond the scope of this paper is required. The MMA-P analysis discussed in Sec. IV B 1 identified four pairs of isomorphic distractors in the module structure of the QMCA. The only item groups that were identified by the MMA-P analysis had correlations greater than a medium effect size for their correlations.

B. Full instrument

In an effort to uncover any underlying structure to the QMCA, we applied MMA-P from Yang *et al.* [13] to all student responses on the full instrument. We require that responses be selected by more than ten test takers, edges in the bootstrapped networks represent correlations that have at least a small effect size ($r < 0.1$) and are significant at a Bonferroni-corrected $\alpha \leq 0.05$, and responses must appear in the same community in 80% of the 1000 iterations.

When MMA-P is applied to the QMCA, twenty modules result, shown in Fig. 3. There are ten modules entirely made up of correct responses and ten modules entirely made up of incorrect responses. In Fig. 3, the correct responses are represented by green nodes and the incorrect responses by pink nodes. The communities reflect the structure of the exam: eight communities include responses to isomorphic items and twelve are

made of up responses from explain pairs. There are no communities that connect items that are not directly related in one of those two ways.

1. Isomorphic items

Since the isomorphic items were described in Sec. IV A, we will not have a detailed discussion of these modules here. There were eight modules that included responses from two isomorphic items, four were correct and four were incorrect, shown in Table III. The pair of modules Iso3 and Iso7 are a correct pair and an incorrect pair of responses, respectively, to the same isomorphic item. Modules Iso4 and Iso8 are likewise a correct and incorrect pair for a different isomorphic question. These isomorphic items have a highly attractive incorrect response that students select in both the spins and wave function contexts.

The correct answer pair for the items in Iso5 (1C-22C) is the only pair of correct responses with a significant partial correlation in Table I not to appear in the MMA-P results. These questions are shown in Fig. 2. In this pairing, the incorrect answer option D in both the spins and wave functions version cluster, but the correct answers (option C) do not cluster. Instead, the correct answer to the spins version clusters with another spins question and its isomorphic pair (this module is labeled ISO 1 in Fig. 3). Iso5 and Iso6 involve items 1 and 22 and 2 and 23, respectively, where items 1 and 2 and items 22 and 23 have common stems.

Item 22, in the spins context, had the second highest number of correct responses (88% percent of students answered it correctly) of all items on the instrument and Items 1, 2, and 23 (responses to the latter two make up Iso1 and Iso6) were all in the top nine of items most often answered correctly. Even so, 1C and 22C were selected together in proportion to the overall score, whereas 2C and 23C were selected together more often than the student's

FIG. 3. Communities of responses commonly selected together by students identified by the MMA-P analysis of the QMCA instrument, including the correct (green nodes) and incorrect (pink nodes) responses. Edges represent the partial correlation, given by the label, between the responses and only edges with correlation of small effect size or greater are shown. There are no intercommunity edges; each set connects nodes in a single community. Cyan edges connect responses to isomorphic items, blue edges connect responses to explain pairs, and black edges are none of these. The thickness of the edges is proportional to the number of times in all the bootstrapped networks that the two responses were chosen together. The communities consist of nodes that clustered together in at least 80% of the bootstrapped networks.

overall scores would suggest. More students in the sample were able to correctly find probabilities of states in both the wave and spins context than were able to identify the possible energy values of those states.

There was one instance of an isomorphic pair whose correct and most common incorrect responses were found to be correlated, but only the correct responses appear as a module in Fig. 3 (Iso2). This is because pairs with partial

TABLE III. Modules identified in the QMCA that consist of isomorphic items. The modules are labeled as in Fig. 3. The expert interpretation of each module is also given.

Module	Type	Expert concept labels for each question	Responses
Iso1 (Iso5) (Iso6)	Correct (Incorrect)	Questions about the energy (spin) of a superposition state and distinguishing the probability from the eigenvalue.	2C, 22C (1D, 22D) (2E, 23E)
Iso2	Correct	Question asking for the state of the system after an energy (spin) measurement.	2C, 23C
Iso3 (Iso7)	Correct (Incorrect)	The state of the system after measurement IS NOT (IS) equal to the corresponding operator acting on the state.	7B, 28B 7A, 28A
Iso4 (Iso8)	Correct (Incorrect)	Question about the effect of a phase of the probability amplitude on the probability.	9A, 30A (9C, 30C)

TABLE IV. Modules identified in the QMCA that consist of explain pairs. The modules are labeled as in Fig. 3. The expert interpretation of each module is also given.

Module	Type	Expert concept labels for each question	Questions
EP1 (EP7)	Correct (Incorrect)	The probability of measuring a particular energy in a superposition of energy eigenstates in repeated measurements is independent (dependent) of time	Q4, Q5
EP2 (EP8)	Correct (Incorrect)	Both the real and imaginary parts are (only the real part is) needed for an arbitrary position measurement on a time-dependent wave function.	Q13, Q14
EP3 (EP9)	Correct (Incorrect)	The position probability density of a superposition of energy eigenstates in a one-dimensional infinite potential well varies with time (are time independent).	Q18, Q19
EP4 (EP10)	Correct (Incorrect)	A system in an eigenstate of an operator will stay (not necessarily stay) in that eigenstate until disturbed by measurement	Q20, Q21
EP5 (EP11)	Correct (Incorrect)	The probability of measuring the spin of a particle in a magnetic field will not depend on time if the spin operator commutes with the Hamiltonian (will always depend on time because it precesses around the magnetic field).	Q25, Q26
EP6 (EP12)	Correct (Incorrect)	The time evolution of a system is (NOT) determined by its Hamiltonian	Q37, Q38

correlation values less than 0.32 do not meet the criteria of the MMA-P modules, even when statistically significant. Since this threshold is a larger than medium effect size (0.30 for a partial correlation), MMA-P is a stricter measure of connected responses than the partial correlation alone.

2. Explain pairs

All of the explain pairs on the QMCA were independently found by MMA-P, appearing both as correct-only and incorrect-only modules. These modules are described in Table IV. Modules EP8 and EP12 have the incorrect answer on true or false items Q13 and Q37 connected to two possible explanations, meaning that these pairs each have two attractive incorrect responses. Both of these modules contain questions about the time evolution of a quantum state, suggesting that several incorrect ideas regarding time dependence are common amongst students. The response to Q19 in module EP9 includes an incorrect explanation about the time variation of the position probability density for a superposition of energy eigenstates of a particle in an infinite square well. Together, modules EP8, EP9, and EP12 (all incorrect) are consistent with treating solutions to the time-independent Schrödinger equation as the full solution to the time-dependent Schrödinger equation. In other words, once separation of variables is performed on the TDSE, the time-dependent equation is ignored and the TISE (also known as the energy eigenequation) is treated as a complete description of all states.

V. DISCUSSION AND CONCLUSIONS

Statistical analysis and reliability testing are important parts of the design, development, and validation of concept assessments. In this paper we have utilized MMA-P on

version 6.2.2 of the QMCA to find communities of student responses. Our goals in this work were to answer two research questions: to determine correlations in student responses to questions designed to be isomorphic in the spins and wave functions context, and to investigate the structure of the QMCA by exploring students' response patterns that emerge from MMA-P.

To investigate the correlations between students' responses to nine paired (isomorphic) questions in the spatial wave function context and the spin context, we measured the partial correlations between the responses to an item in each context, controlling for overall score on the QMCA. Of the nine pairs, we found that five of these were highly correlated, in that students responded to both questions in the pair consistently (either correctly or incorrectly). These five paired questions were about measurements where the state is written in the basis of the operator being measured, thus implying that finding probabilities and identifying the state after measurement are transferable between contexts. These pairs also appeared in the MMA-P modules. That leaves four questions that did not correlate between the spins and wave functions contexts, indicating that students did not choose a consistent pair of responses to these questions. These questions were related to noncommuting operators, the energy eigenstates, and the time evolution of the system. The fact that these questions do not correlate is informative, but not entirely surprising, as the spin and position operators are very different both mathematically and conceptually. These questions may also provide some insight into the different ways experts and students view these concepts [34–36]. An expert might see a question about the time evolution of a quantum system and think abstractly about the energy eigenstates of the system and whether the state of interest is

a stationary state without spending much time thinking about the physical system, whereas a student may be focused on the details of the problem, like the fact that the particle is a spin in a magnetic field instead of a spinless particle in a square well potential. In addition, repeated energy and position measurements appear to not be viewed the same by students as repeated measurements about spin in different directions. This may be similar to the expert or student distinction in introductory physics where an expert may look at a problem and recognize it as one where energy is conserved, where a student may see the same question and focus on the it being a problem where a box slides down an inclined plane. The nature of this connection warrants further study.

Our second research question is more open ended, as we were hoping to explore any potential structure to the QMCA that may help guide both future investigations of student learning as well as curriculum and assessment development. In the initial development of the QMCA, the creators categorized the questions into five groups based on course learning goals and we were interested to see if that categorization appears in the module analysis. What we found was that MMA did not reveal a large-scale structure to student responses on the QMCA, either by the initial categorizations by the survey developers or otherwise. Instead, what we found were many pairs of questions that were correlated. This is similar to the structure found in the Force Concept Inventory (FCI) [13,14] and unlike that found in the Force and Motion Conceptual Evaluation (FMCE) [15], which had larger communities. The QMCA and the FCI are similar in structure; both were designed to present students with commonly held incorrect ideas. The FMCE gives all possible responses, regardless of how common the idea is among students, e.g., all combinations of direction of acceleration and velocity vectors to label a cart moving up a ramp. The QMCA cannot find gradations within a given community of student responses, compared to an instrument, such as the FMCE, where the communities contained numerous items. On the other hand, the lack of communities that mix correct and incorrect responses may indicate that few items are malfunctioning, since one potential reason for correct and incorrect responses appearing together in the same community is that the question is not being interpreted by students in the way an expert would expect.

In the absence of any large-scale structure, we focus on smaller structural categorizations. Items that form explain pairs are often the basis for modules. Explain pairs are generally two questions where the first asks a question with a yes or no (or true or false) answer and then the second question is an extension of the first, often by asking for students to select the best explanation to the previous question. We found that each instance of these explain pair questions on the QMCA are correlated, with both correct and incorrect answer correlations. Incorrect answers

to the first question commonly correlate with one or two distractor explanations in the second question. When designing the QMCA, the explain pair question type was used to replicate as closely as possible an open-ended question with a closed-response and the fact that both the correct and incorrect response pairing cluster independently is validation that these pairs are capturing student ideas.

On the QMCA there are two questions that were designed to test the idea that making a measurement on a quantum system is not the same as acting an operator on the quantum state [37]. These questions were asked in both the spatial wave function context and the spin context. The question stem provided students with a quantum state in a superposition of energy eigenstates (in the wave functions context) or of S_z eigenstates. The first question in the pair asked students what the state of the system would be after energy (or S_z) were measured. The second was a true or false question for the statement: immediately after an energy measurement the system is in the state $H\psi$. In neither context (wave functions nor spins) was there a significant partial correlation between the responses to these two questions. However, both questions are correlated to their “isomorphic equivalent.” The idea that acting an operator on a state does not constitute a measurement has been studied in the past and is known to be difficult for students. The fact that these QMCA questions are not correlated supports those findings.

The QMCA instrument includes common stem questions that provide a single context for a sequence of questions that may or may not be related by concept. The three QMCA items (34–36) regarding ColorTron form an example of common stem questions. The set consists of text at the top that describes a quantum system involving measurements of whimsical quantum mechanical observables “color” and “size” and their related eigenvalues. The subsequent three questions inquire about results of (i) a repeated measurement of color; (ii) a normalized representation of a given state in size basis; and (iii) repeated measurements of commuting operators of color and size.

In our previous research, the ColorTron common stem often appeared as one of the most difficult sets of questions [7]. In a previous MMA analysis of the FCI [14], the results suggested that common stem items should not be scored independently. However, on the QMCA none of the common stem question sets appears as a module in the MMA analysis. In fact, there are no significant correlations among students’ responses to any common stem questions on the QMCA. This suggests that in the analysis of diagnostic instruments, one should not by default assume a correlation among students’ responses to common stem questions.

The QMCA is the first concept inventory intended for use in an upper division course to be analyzed with MMA-P. Additionally, the QMCA is only given as a

post-test, unlike the concept inventories which have been studied using MMA-P thus far. As is often the case, upper division courses have fewer students relative to introductory courses, therefore this analysis uses a smaller sample than all but the original module analysis paper, which had roughly half the number of exams that are included here [19]. Because of this, the responses from the different institutions which provided their QMCA results for this study were aggregated into a single sample. The sample sizes from each institution were not equal, though the bootstrapping should preserve the proportion of students from each institution in the MMA-P analysis. It is possible that the students at these different institutions may have different patterns of responses; in this case the MMA-P analysis would find the most common correct and incorrect response communities for this group of students, but perhaps not the most common patterns at any particular institution. Differing incentives and delivery methods (paper vs online) may also affect the seriousness with which students take the exam, though there is evidence that the delivery method does not have a large effect on student effort [18]. The dataset presented here came only from courses using interactive engagement pedagogy and following a spins-first instructional paradigm. Since the majority of quantum courses in the U.S. are taught using

positions-first curriculum, it would be useful and interesting to expand studies of this type to include more diverse student populations.

We have performed the first broad-scale statistical analysis of the QMCA going beyond simple classical test analysis. The novel clustering and lack of statistical evidence for the anticipated topical question categories indicates that MMA can be a productive quantitative method to analyze conceptual surveys. We found that there was no large-scale structure revealed by the MMA-P analysis and instead found pairs of questions that were highly correlated, namely, some of the isomorphic pairs, and explain pair questions. The results also show that approximately half of the questions designed to be isomorphic had strong correlations between student responses in the spins and position contexts, while the other half were not correlated. These results provide additional insights into the structure of the QMCA and its suitability to assess student learning in different instructional contexts.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Grants No. DUE 1626280, No. 1626594, and No. 1626482.

-
- [1] G. Zhu and C. Singh, Surveying students' understanding of quantum mechanics in one spatial dimension, *Am. J. Phys.* **80**, 252 (2012).
 - [2] G. Zhu and C. Singh, Improving students' understanding of quantum measurement. I. Investigation of difficulties, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010117 (2012).
 - [3] J. Falk, Developing a quantum mechanics concept inventory, Master's thesis, Uppsala University, Uppsala, Sweden, 2004.
 - [4] E. Cataloglu and R. W. Robinett, Testing the development of student conceptual and visualization understanding in quantum mechanics through the undergraduate career, *Am. J. Phys.* **70**, 238 (2002).
 - [5] S. B. McKagan, K. K. Perkins, and C. E. Wieman, Design and validation of the Quantum Mechanics Conceptual Survey, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020121 (2010).
 - [6] J. Stewart, C. Zabriskie, S. DeVore, and G. Stewart, Multidimensional item response theory and the Force Concept Inventory, *Phys. Rev. Phys. Educ. Res.* **14**, 010137 (2018).
 - [7] H. R. Sadaghiani and S. J. Pollock, Quantum mechanics concept assessment: Development and validation study, *Phys. Rev. ST Phys. Educ. Res.* **11**, 010110 (2015).
 - [8] B. Ambrose, Investigation of student understanding of the wavelike properties of light and matter, Ph.D. thesis, University of Washington, Seattle, Washington, 1999.
 - [9] H. Sadaghiani, Conceptual and mathematical barriers to students learning quantum mechanics, Ph.D. thesis, The Ohio State University, Columbus, Ohio, 2005.
 - [10] A. Crouse, Research on student understanding of quantum mechanics as a guide for improving instruction, Ph.D. thesis, University of Washington, Seattle, Washington, 2007.
 - [11] C. Singh, Student understanding of quantum mechanics at the beginning of graduate instruction, *Am. J. Phys.* **76**, 277 (2008).
 - [12] S. Pollock, G. Passante, and H. Sadaghiani, Adaptable curricular exercises for quantum mechanics, <https://www.physport.org/curricula/ACEQM/>.
 - [13] J. Yang, J. Wells, R. Henderson, E. Christman, G. Stewart, and J. Stewart, Extending modified module analysis to include correct responses: Analysis of the Force Concept Inventory, *Phys. Rev. Phys. Educ. Res.* **16**, 010124 (2020).
 - [14] J. Wells, R. Henderson, J. Stewart, G. Stewart, J. Yang, and A. Traxler, Exploring the structure of misconceptions in the Force Concept Inventory with modified module analysis, *Phys. Rev. Phys. Educ. Res.* **15**, 020122 (2019).
 - [15] J. Wells, R. Henderson, A. Traxler, P. Miller, and J. Stewart, Exploring the structure of misconceptions in the Force and Motion Conceptual Evaluation with modified module analysis, *Phys. Rev. Phys. Educ. Res.* **16**, 010121 (2020).
 - [16] S. Pollock, H. Sadaghiani, A. Quaal, and G. Passante, Designing, validating, and contrasting conceptual quantum

- mechanics questions for spin states and spatial wave functions, in *Proceedings of the 2018 Physics Education Research Conference, Washington, DC* (AIP, New York, 2018), <https://doi.org/10.1119/perc.2018.pr.Pollock>.
- [17] D. H. McIntyre, *Quantum Mechanics* (Pearson Education, Inc., San Francisco, CA, 2012).
- [18] B. R. Wilcox and S. J. Pollock, Investigating students' behavior and performance in online conceptual assessment, *Phys. Rev. Phys. Educ. Res.* **15**, 020145 (2019).
- [19] E. Brewe, J. Bruun, and I. G. Bearden, Using module analysis for multiple choice responses: A new method applied to Force Concept Inventory data, *Phys. Rev. Phys. Educ. Res.* **12**, 020131 (2016).
- [20] C. Wheatley, J. Wells, R. Henderson, and J. Stewart, Applying module analysis to the Conceptual Survey of Electricity and Magnetism, *Phys. Rev. Phys. Educ. Res.* **17**, 010102 (2021).
- [21] D. Huffman and P. Heller, What does the force concept inventory actually measure? *Phys. Teach.* **33**, 138 (1995).
- [22] T. F. Scott, D. Schumayer, and A. R. Gray, Exploratory factor analysis of a Force Concept Inventory data set, *Phys. Rev. ST Phys. Educ. Res.* **8**, 020105 (2012).
- [23] A. Lichtenberger, C. Wagner, S. I. Hofer, E. Stern, and A. Vaterlaus, Validation and structural analysis of the kinematics concept test, *Phys. Rev. Phys. Educ. Res.* **13**, 010115 (2017).
- [24] P. Eaton, B. Frank, K. Johnson, and S. Willoughby, Comparing exploratory factor models of the Brief Electricity and Magnetism Assessment and the Conceptual Survey of Electricity and Magnetism, *Phys. Rev. Phys. Educ. Res.* **15**, 020133 (2019).
- [25] R Core Team, R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2017).
- [26] RStudio Team, *RStudio: Integrated Development Environment for R*, (RStudio, PBC, Boston, MA, 2020).
- [27] S. Fortunato and D. Hric, Community detection in networks: A user guide, *Phys. Rep.* **659**, 1 (2016).
- [28] A. Clauset, M. E. J. Newman, and C. Moore, Finding community structure in very large networks, *Phys. Rev. E* **70**, 066111 (2004).
- [29] M. E. J. Newman, Modularity and community structure in networks, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8577 (2006).
- [30] G. Csardi and T. Nepusz, The igraph software package for complex network research, *InterJournal Complex Systems* **1695**, 1 (2006).
- [31] H. Wickham, R. François, L. Henry, and K. Müller, *dplyr: A Grammar of Data Manipulation* (2018), R package version 0.7.7.
- [32] W. Revelle, *psych: Procedures for Psychological, Psychometric, and Personality Research* (Northwestern University, Evanston, Illinois, 2018), R package version 1.8.12.
- [33] R. E. Scherr, Modeling student thinking: An example from special relativity, *Am. J. Phys.* **75**, 272 (2007).
- [34] J. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, Expert and novice performance in solving physics problems, *Science* **208**, 1335 (1980).
- [35] M. T. H. Chi, P. J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, *Cogn. Sci.* **5**, 121 (1981).
- [36] S.-Y. Lin and C. Singh, European Physical Society, find out more Categorization of quantum mechanics problems by professors and students, *Eur. J. Phys.* **31**, 57 (2010).
- [37] E. Gire and C. Manogue, Making sense of quantum operators, eigenstates and quantum measurements, *AIP Conf. Proc.* **1413**, 195 (2012).