Using the Rasch model to analyze the test of understanding of vectors

Sušac, Ana; Planinić, Maja; Klemenčić, Damjan; Milin Šipuš, Željka

Using the Rasch model to analyze the test of understanding of vectors

Ana Susac,[1,*] Maja Planinic,[2] Damjan Klemencic,[2,3] and Zeljka Milin Sipus[3]

[1]*Department of Applied Physics, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia*
[2]*Department of Physics, Faculty of Science, University of Zagreb, Bijenicka 32, 10000 Zagreb, Croatia*
[3]*Department of Mathematics, Faculty of Science, University of Zagreb, Bijenicka 30, 10000 Zagreb, Croatia*

The test of understanding of vectors (TUV) is a multiple-choice test that was recently developed to assess student understanding of vector concepts required in introductory physics courses at the university level. In this study, the TUV was administered to 889 first-year students at the University of Zagreb. The Rasch model was used to evaluate the functioning of the TUV, and to determine the difficulty scale of vector concepts evaluated by this assessment instrument. The results of the Rasch analysis showed a good functioning of test items. However, the TUV was not well targeted to our sample of first-year engineering and science students in introductory physics courses. About half of the students had abilities outside the range of item difficulties, indicating that the TUV would benefit from including more difficult items that would help to better estimate the abilities of these students. Further analysis of the data from subpopulations in our sample showed that the TUV can be well targeted for certain student populations. Students' scores suggest that the most difficult vector concept tested by the TUV is the unit vector, followed by the cross product, subtraction of vectors, dot product, and direction of a vector.

I. INTRODUCTION

Many physics quantities are vectors, and it is essential for students to acquire a good understanding of vector graphical and algebraic properties. Physics education research (PER) studies revealed a number of student difficulties with both vectors in general and with vectors in specific physics contexts [1–10]. Based on these previous findings and their own meticulous studies [11–16], Barniol and Zavala developed the test of understanding of vectors [17].

The PER community has adopted different approaches to data analysis of multiple-choice questions [18]. Widely used physics diagnostic tests are often evaluated in multiple ways, e.g., using classical test theory, item response theory, factor analysis, or Rasch analysis [19–24]. In the case of the test of understanding of vectors (TUV), the classical test theory was used for evaluations during the development stage and for the analysis of the final version [17]. The three-parameter logistic model of item response theory

(IRT) and item response curves technique were employed to analyze the TUV when it was released [25].

The Rasch model is another useful tool for the evaluation of tests that are intended to be used as assessment instruments [18]. As the Rasch model-based analysis of the TUV has not yet been performed, we decided to reevaluate the TUV using that approach.

In this paper we aimed to answer the following research questions:
 (i) How does the TUV function on a sample of first-year engineering and science students?
 (ii) What difficulty scale of vector concepts is suggested by students' scores on the TUV?

We performed the Rasch model-based analysis of the TUV to address these research questions.

II. METHODS

A. Participants

The study included 889 undergraduate first-year engineering and science students from the University of Zagreb. A detailed description of the sample is given in the Supplemental Material [26].

B. Data collection

The test of understanding of vectors [17] was translated in Croatian and validated by two university lecturers in

*Corresponding author.
ana.susac@fer.hr

physics and one university lecturer in mathematics. The test was administered to the participants during their second semester at the university.

C. Data analysis

We used the Rasch model to analyze the collected data. The Rasch model is an important psychometric tool when conducting science education research utilizing multiple-choice tests [27]. For an introduction to Rasch analysis see, for example, Ref. [28]. We used the Winsteps software [29] to conduct the Rasch analysis.

III. RESULTS AND DISCUSSION

A. Analysis of test structure and functioning

The Rasch analysis of the collected data showed very high item reliability (0.99), which indicates replicability of item order according to difficulty if the test were administered to another similar sample of students. The person reliability was rather low (0.60), which indicates a not very reliable person order if a similar test was administered to the same students. The person reliability reports the reproducibility of the person measure order obtained through Rasch analysis. This index can range from 0 to 1, but its minimum meaningful value is 0.5, whereas the lowest person reliability for any decision making about students' abilities (e.g., discernment between high and low performers) is 0.8. The most effective way to increase the person reliability is to increase the number of test items [30]. The Cronbach alpha was 0.79, which can be considered satisfactory. The obtained value corresponds to the value of the Kuder-Richardson reliability index (0.78) reported by Barniol and Zavala for their sample of students [17].

Figure 1 shows the distributions of student abilities and item difficulties on the same logit scale. One logit is the distance along the scale that increases the probability of observing the event specified in the measurement model by a factor of 2.718 (the base of the natural logarithm e). The average item difficulty is set at zero. More able students and more difficult items have a more positive value of Rasch measure. Figure 1 reveals that the mean student ability is about 2 logit above the mean item difficulty, which indicates a rather poor test targeting. The test is too easy for our sample of first-year university students. A significant fraction of students, 11.5% (102 students), solved all test items correctly, and an additional 13.3% (118 students) solved 19 test items correctly. About half of the students had abilities outside the range of item difficulties. The test does not contain items that would help to better estimate the abilities of these students. Consequently, the person reliability is quite low, as reported above.

If the test is meant to be well targeted, then the distribution of item difficulties should be aligned with the distribution of student abilities. Ideally, items should be
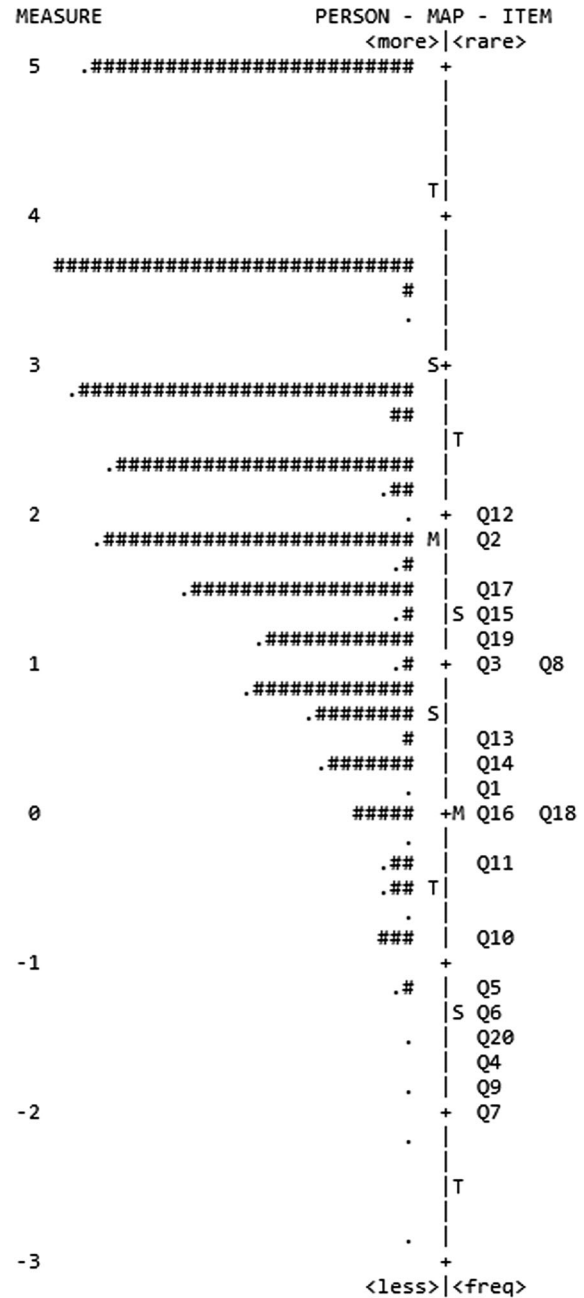

FIG. 1. The person-item map shows the distribution of student abilities on the left-hand side and the distribution of item difficulties on the right-hand side displayed on the same logit scale. Each "hash" represents four students and each "dot" represents one to three students. The TUV items are labeled as Q1–Q20. M denotes the mean of each distribution; S denotes 1 standard deviation, and T 2 standard deviations from the mean.

distributed more or less evenly along the whole range of student abilities, since student ability is best evaluated by items in the ± 1 logit interval around the ability value. For the sample of students in this study, the TUV has enough easy items, but it does not have enough difficult items. There are no items centered on students who are in the

ability range of 2–5 logit. These results indicate that the TUV would benefit from including more difficult items.

To evaluate specificities of different student populations, we conducted the Rasch analysis of the data collected from subpopulations in our sample and the obtained person-item maps are shown in the Supplemental Material [26]. The results show that the TUV had good targeting only for the subsample of students from the Faculty of Chemical Engineering and Technology.

B. Analysis of test items

Furthermore, we examined the fit of the items with the Rasch model by calculating infit and outfit mean square (MNSQ) residuals and standardized Z scores for each item (Table I). Typically, items are considered to have acceptable fit if both infit and outfit MNSQ are between 0.7 and 1.3, and Z scores are between -2 and 2 [28], but items with MNSQ in the range 0.5–1.5 will still be productive for measurement [30]. High infit values of MNSQ and Z scores indicate that students do not respond in the expected way to items whose difficulties correspond to their ability. High outfit values of MNSQ and Z scores usually indicate outliers (e.g., a student of lower ability answers correctly on a very difficult test item). Low infit and outfit values of MNSQ and Z score indicate overfit to the Rasch model, i.e., the data are more predictable than the model expects, and the items therefore do not provide much new information, but they do not degrade the measurement.

TABLE I. Item difficulty measured in logits, Rasch standard error, infit and outfit MNSQ (mean square residuals) and Z scores, and point-measure correlation for each TUV item.

Item	Measure	SE	Infit MNSQ	Infit Z	Outfit MNSQ	Outfit Z	Correlation
Q1	0.23	0.09	1.11	2.0	1.23	2.2	0.37
Q2	1.89	0.08	0.90	−3.1	0.89	−2.1	0.60
Q3	1.06	0.08	0.90	−2.8	0.86	−2.5	0.56
Q4	−1.63	0.16	0.99	0.0	0.77	−0.8	0.27
Q5	−1.18	0.14	1.01	0.1	0.96	−0.1	0.30
Q6	−1.40	0.15	1.01	0.2	0.86	−0.5	0.28
Q7	−2.05	0.20	0.92	−0.4	0.46	−1.8	0.28
Q8	0.94	0.08	1.01	0.2	1.06	1.0	0.49
Q9	−1.84	0.18	0.90	−0.7	1.00	0.1	0.29
Q10	−0.79	0.12	1.05	0.6	1.31	1.6	0.29
Q11	−0.37	0.11	0.91	−1.2	0.75	−1.9	0.44
Q12	1.93	0.08	1.02	0.5	1.27	4.6	0.54
Q13	0.53	0.09	0.94	−1.3	0.89	−1.3	0.50
Q14	0.27	0.09	1.24	4.3	1.42	3.8	0.31
Q15	1.32	0.08	0.91	−2.5	0.87	−2.4	0.57
Q16	−0.06	0.10	0.91	−1.5	0.80	−1.8	0.46
Q17	1.56	0.08	1.19	5.2	1.27	4.8	0.44
Q18	−0.01	0.10	0.93	−1.2	0.86	−1.2	0.46
Q19	1.13	0.08	1.03	0.8	1.03	0.5	0.50
Q20	−1.54	0.16	1.01	0.1	1.00	0.1	0.27

Table I shows that most items fit well with the model. The item that is problematic to a certain extent, because of a larger misfit, is item 14. Rather high outfit values of its MNSQ and Z score were largely caused by some students of higher ability who unexpectedly failed on this test item. This agrees with the results from the previous study by Rakkapao *et al.* [25], where rather low discrimination power was found for this item. Item 14 refers to calculation of the x component of a vector when the angle is measured from the y axis. The majority (70%) of students who failed on this item chose the distractor in which the sine function was replaced by cosine. This might be caused by students' lack of attention (angle was measured from the y axis) or their difficulties with trigonometric functions. Overall, item 14 should be further examined and possibly revised in a future version of the TUV. Items 17 and 12 also show high values of Z score, but their MNSQ values are acceptable. Since Z scores are too sensitive for large samples [30], items 17 and 12 can be considered productive for measurement.

The Rasch model assumes unidimensionality, i.e., the existence of a single underlying measurement construct. The point-measure correlations can help identify the existence of the construct. Positive point-measure correlation of an item indicates that the item is in line with the measured construct. The size of correlations shows how much the items contribute to the construct. Table I shows that the point-measure correlations of all items are positive, which suggests that all items measure the underlying construct (understanding of vectors).

C. Difficulty scale of vector concepts

The TUV is developed to test ten vector concepts used in the introductory physics courses [17]. Each vector concept was tested by one to three TUV items. To compare the difficulties of the TUV items related to each vector concept, we calculated their average values and their uncertainties (Fig. 2).

The most difficult vector concept appears to be the unit vector. This concept was tested by only one test item (item 2). This finding is consistent with the results from the previous studies [17,25]. For the group of Thai students, item 2 was the most difficult TUV item, while it was the third most difficult item for the group of Mexican students. The most popular distractor in item 2 in all three studies (see Refs. [17,25] and the present study) was the choice in which the x and y components were unit vectors. Barniol and Zavala [14] and Rakkapao *et al.* [25] found that the students who have chosen this answer believed that this vector had a magnitude 1. So, we might conclude that many students who failed to select the correct answer on item 2 probably knew the definition of unit vector but they did not correctly use the notion of vector magnitude and vector decomposition. To better examine student understanding of
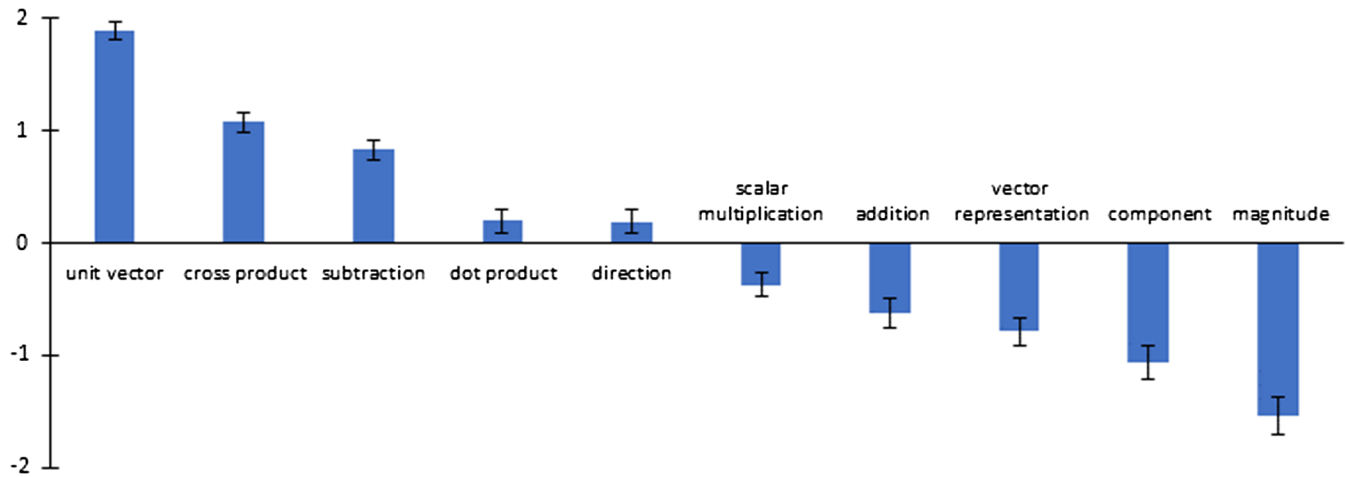
FIG. 2.   Average difficulties of the vector concepts, evaluated by the TUV, measured in logits.

the unit vector, it would be useful to have more test items on this vector concept.

The next two concepts according to difficulty are cross product and subtraction of vectors. Two items on cross product were above the average difficulty, and one item was close to the average. Item 12 (geometric interpretation of the cross product as a perpendicular vector satisfying right-hand rule) and item 15 (calculation of the cross product of vectors written in unit-vector notation) were among the most difficult items, whereas item 18 (identifying the correct formula for cross product magnitude) was considerably easier. Both items 12 and 15 refer to the direction of the vector product, and the most frequent incorrect answers on both items were vectors with the opposite direction to the correct vector, which possibly indicates the misapplication of the right-hand rule. This finding is in agreement with the previous studies on student difficulties with cross product direction [13,16,31].

Subtraction of vectors is significantly more difficult than the addition of vectors, and that was corroborated by the previous reports [17,25]. For our sample of students, subtraction of vectors in one dimension (item 19) was more difficult than subtraction in two dimensions (item 13). Similar result was reported for the group of Thai students [25], while Mexican students had more difficulties with subtraction of vectors in two dimensions [17]. Heckler and Scaife [32] found that many student difficulties with simple vector addition and subtraction lie with the arrow representation itself, so they suggested the introduction of other representation (e.g., unit vectors ijk notation) together with the arrow notation.

The last two vector concepts that were more difficult than the average are dot product and vector direction. Student understanding of the vector direction was tested by item 5 (choosing a vector with the same direction as the given vector from among several options) that was easier than the average, and item 17 (calculation of direction of a

vector written in unit-vector notation) that was the third most difficult item in the test (Fig. 1). Furthermore, the dot product was easier than the cross product for students in this study, similar to the results in the previous study by Rakkapao *et al.* [25]. Barniol and Zavala found the opposite; for their sample of students, the cross product appeared easier than the dot product [17]. This might be caused by the fact that they were tested after the course on electricity and magnetism in which the cross product of vectors is often used. Anyway, the dot product is among the more demanding vector concepts in all three studies. For the sample of students in our study, item 6 (identifying correct formula for a dot product) was significantly easier than items 3 (geometric interpretation of a dot product as a projection) and 8 (calculation of a dot product of vectors written in unit-vector notation). This suggests that knowing the formula is not enough for problem solving [33,34].

The remaining vector concepts (scalar multiplication of a vector, vector addition, graphic representation of a vector, vector components, magnitude of a vector) were below average difficulty, and did not pose a problem for most of the students.

IV. CONCLUSION

The Rasch analysis of the test of understanding of vectors showed rather good functioning of test items. Only item 14 deserves further inspection and possibly revision. However, the TUV was not well targeted to our sample of first-year engineering and science students in introductory physics courses. The mean of the distribution of item difficulties was about two logit below the mean of the distribution of student abilities indicating that the test was too easy for our sample. The lack of more difficult test items resulted in low person reliability. Nevertheless, the analysis of the data from subpopulations in our sample showed that the TUV can be well targeted for certain

student populations [26]. Furthermore, its functioning might depend on educational systems and teaching of vector concepts. For example, our sample of first-year engineering and science students in Croatia had better results on the TUV than Mexican and Thai students from the previous studies [17,25]. The possible reason for the observed difference might be that Croatian students learned about some of the vector concepts tested in this study already in high school.

One could argue that the understanding of vector concepts is the essential prerequisite for attending introductory physics courses and that high student scores on the TUV are thus expected. In that case the TUV could be applied as a pass or fail assessment instrument. High scores on the TUV would indicate that most students acquired good understanding of vectors, and that they do not have difficulties with vector quantities and operations with vectors. Unfortunately, from our teaching practice, we could not confirm that is the case. Students in our introductory courses beyond mechanics still struggle with some vector concepts, such as cross and dot products. Thus, taking into account the results of the Rasch analysis and our insight in student difficulties with vector quantities in physics, we suggest adding more difficult items in the TUV.

Besides adding more items on the vector concepts already included in the TUV, but with only one item (e.g., unit vector), it might be beneficial to include some other vector problems often used in introductory physics courses (e.g., vector decomposition in nonorthogonal directions, adding or subtracting vectors that do not have the same initial point, finding the difference of equally long vectors with opposite directions, or, when given a resulting vector of a cross product and one factor, finding the possible directions of the unknown factor). These are the examples of the less-standard vector problems encountered in introductory physics, which may present difficulty for students. It is our experience that students indeed do have difficulties applying their knowledge of vectors to such situations, and that is something that should be detectable by TUV and similar diagnostic instruments.

According to the results of the Rasch analysis, it seems that the most difficult vector concepts tested by the TUV is the unit vector, followed by the cross product, subtraction of vectors, dot product, and the direction of a vector. The order of vector concepts by difficulty might help instructors in introductory physics courses to put more emphasis on these concepts in their teaching in order to help students to overcome the observed difficulties.

[1] J. M. Aguirre and G. Rankin, College students' conceptions about vector kinematics, Phys. Educ. **24**, 290 (1989).

[2] R. D. Knight, The vector knowledge of beginning physics students, Phys. Teach. **33**, 74 (1995).

[3] N. Nguyen and D. E. Meltzer, Initial understanding of vector concepts among students in introductory physics courses, Am. J. Phys. **71**, 630 (2003).

[4] S. Flores, S. E. Kanim, and C. H. Kautz, Student use of vectors in introductory mechanics, Am. J. Phys. **72**, 460 (2004).

[5] P. S. Shaffer and L. C. McDermott, A research-based approach to improving student understanding of the vector nature of kinematical concepts, Am. J. Phys. **73**, 921 (2005).

[6] J. Van Deventer and M. C. Wittmann, Comparing student use of mathematical and physical vector representations, AIP Conf. Proc. **951**, 208 (2007).

[7] T. Wang and E. C. Sayre, Maximum likelihood estimation (MLE) of students' understanding of vector subtraction, AIP Conf. Proc. **1289**, 329 (2010).

[8] J. M. Hawkins, J. R. Thompson, and M. C. Wittmann, Students' consistency of graphical vector addition method on 2-D vector addition tasks, AIP Conf. Proc. **1179**, 161 (2009).

[9] J. M. Hawkins, J. R. Thompson, M. C. Wittmann, E. C. Sayre, and B. W. Frank, Students' responses to different

representations of a vector addition question, AIP Conf. Proc. **1289**, 165 (2010).

[10] B. E. Hinrichs, Writing position vectors in 3-d space: A student difficulty with spherical unit vectors in intermediate E&M, AIP Conf. Proc. **1289**, 173 (2010).

[11] P. Barniol and G. Zavala, Investigation of students' preconceptions and difficulties with the vector direction concept at a Mexican university, AIP Conf. Proc. **1179**, 85 (2009).

[12] P. Barniol and G. Zavala, Vector addition: Effect of the context and position of the vectors, AIP Conf. Proc. **1289**, 73 (2010).

[13] G. Zavala and P. Barniol, Students' understanding of the concepts of vector components and vector products, AIP Conf. Proc. **1289**, 341 (2010).

[14] P. Barniol and G. Zavala, Students' difficulties with unit vectors and scalar multiplication of a vector, AIP Conf. Proc. **1413**, 115 (2012).

[15] G. Zavala and P. Barniol, Students' understanding of dot product as a projection in no-context, work and electric flux problems, AIP Conf. Proc. **1513**, 438 (2013).

[16] P. Barniol, G. Zavala, and C. Hinojosa, Students' difficulties in interpreting the torque vector in a physical situation, AIP Conf. Proc. **1513**, 58 (2013).

[17] P. Barniol and G. Zavala, Test of understanding of vectors: A reliable multiple-choice vector concept test, Phys. Rev. ST Phys. Educ. Res. **10**, 010121 (2014).

[18] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, Phys. Rev. ST Phys. Educ. Res. **5**, 020103 (2009).

[19] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, Phys. Teach. **30**, 141 (1992).

[20] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, Phys. Rev. ST Phys. Educ. Res. **2**, 010105 (2006).

[21] J. Wang and L. Bao, Analyzing force concept inventory with item response theory, Am. J. Phys. **78**, 1064 (2010).

[22] M. Planinic, L. Ivanjek, and A. Susac, Rasch model based analysis of the Force Concept Inventory, Phys. Rev. ST Phys. Educ. Res. **6**, 010103 (2010).

[23] 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).

[24] L. Ding, Seeking missing pieces in science concept assessments: Reevaluating the brief electricity and magnetism assessment through Rasch analysis, Phys. Rev. ST Phys. Educ. Res. **10**, 010105 (2014).

[25] S. Rakkapao, S. Prasitpong, and K. Arayathanitkul, Analysis test of understanding of vectors with three-parameter logistic model of item response theory and item response curve technique. Phys. Rev. Phys. Educ. Res. **12**, 020135 (2016).

[26] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.14.023101, for detailed description of the sample and the person-item maps for different groups of students.

[27] W. J. Boone and K. Scantlebury, The role of Rasch analysis when conducting science education research utilizing multiple-choice tests, Sci. Educ. **90**, 253 (2006).

[28] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (Lawrence Erlbaum, Mahwah, NJ, 2001).

[29] J. M. Linacre, Winsteps Rasch measurement computer program. Beaverton, Oregon: Winsteps.com (2017).

[30] J. M. Linacre, A User's Guide to WINSTEPS, http://www.winsteps.com.

[31] M. B. Kustusch, Assessing the impact of representational and contextual problem features on student use of right-hand rules, Phys. Rev. Phys. Educ. Res. **12**, 010102 (2016).

[32] A. F. Heckler and T. M. Scaife, Adding and subtracting vectors: The problem with the arrow representation, Phys. Rev. ST Phys. Educ. Res. **11**, 010101 (2015).

[33] L. Ivanjek, A. Susac, M. Planinic, A. Andrasevic, and Z. Milin-Sipus, Student reasoning about graphs in different contexts, Phys. Rev. Phys. Educ. Res. **12**, 010106 (2016).

[34] A. Susac, A. Bubic, E. Kazotti, M. Planinic, and M. Palmovic, Student understanding of graph slope and area under a graph: A comparison of physics and nonphysics students, Phys. Rev. Phys. Educ. Res. **14**, 020109 (2018).