Journal of Turkish Science Education, 2022, *19*(4), 1312-1327.
DOI no: 10.36681/tused.2022.176

Journal of Turkish Science Education
http://www.tused.org
© ISSN: 1304-6020

Development of a Circular Motion Concept Question Item Inventory for Use in Ugandan Science Education

Kent Robert Kirya[1], Kalarattu Kandiyi Mashood[2], Lakhan Lal Yadav[3]

[1]*African Centre of Excellence for Innovative Teaching and Learning Mathematics and Science (ACEITLMS), University of Rwanda - College of Education (UR-CE), Rwanda. kiryakirya038@gmail.com ORCID ID: 0000-0002-9759-6335*

[2]*Homi Bhabha Centre for Science Education, Tata Institute of Fundamental Research, India, ORCID ID: 0000-0002-3408-1553*

[3]*African Centre of Excellence for Innovative Teaching and Learning Mathematics and Science (ACEITLMS), University of Rwanda - College of Education (UR-CE), Rwanda, ORCID ID: 0000-0003-1182-8017*

ABSTRACT
In this study, we administered and evaluated circular motion concept question items with a view to developing an inventory suitable for the Ugandan context. Before administering the circular concept items, six physics experts and ten undergraduate physics students carried out the face and content validation. One hundred eighteen undergraduate students responded to the 42 circular motion concept items. The data were analysed using the classical test theory (CTT) and item response curve (IRC) analyses. We calculated the difficulty level and index of discrimination and gauged the distraction efficiency of items. The IRCs revealed insights that were not evident from those provided by the CTT. Based on the IRCs, the circular concept items are classified into three categories: efficient, moderately efficient, and inefficient. This helped us better evaluate the quality of the items and their appropriateness for the population under consideration. We ended up with 22 circular motion concept question items which we call the circular motion concept inventory (CMCI). This inventory is particularly relevant to Ugandan context and may be useful to other countries in the East African region which share similar syllabi.

RESEARCH ARTICLE

ARTICLE INFORMATION
Received: 30.05.2022
Accepted: 08.11.2022

KEYWORDS:
Physics education research, concept inventory, circular motion concept inventory (CMCI), classical test theory (CTT), item response curves (IRCs).

To cite this article: Kirya, K.R., Mashood, K.K., & Yadav, L.L. (2022). Development of a circular motion concept question item inventory for use in Ugandan Science Education. *Journal of Turkish Science Education, 19*(4), 1312-1327.

Introduction

Since the middle of the 1980s, one of the most significant changes in physics education research (PER), as well as other STEM fields, has been the increasing construction and use of concept inventories (CIs). A CI is a research-based assessment tool that uses multiple-choice questions to assess pupils' conceptual understanding. A well-developed, valid and reliable CI can be used to identify conceptual difficulties and measure changes in conceptual understanding. Before instruction, CIs are frequently utilised as assessment tools to obtain insights into learners' alternative conceptions and to measure the efficacy of instruction. In science, CIs often concentrate on assessing the "conceptual change" or "expert-like thinking" of learners. Different CIs have been created using various methodological strategies, although they frequently use the same process (Adams & Wieman, 2011; Ding & Beichner, 2009; Laverty & Caballero, 2018; Mashood & Singh, 2015; Mashood, 2014). To develop a good concept inventory, construction of good questions, covering major concepts, with appropriate distractors, usually based on common student alternative conceptions, is very important.

Typically, the development of a CI begins with content mapping and the development of several questions using the targeted concepts as a guide. These introductory questions, which are usually open-ended, are asked repeatedly to representatives of the intended audience (learners) during tests, think-aloud interviews, or both. The questions are developed iteratively by considering the learners' responses. Questions that are judged inappropriate are those that either practically all learners correctly answer or that they do not interpret as anticipated. Then, using distractor answer possibilities that correspond to the responses tendered, multiple-choice questions are created from their inaccurate answers, together with the correct answer. It is typical for the marking criteria for open-ended tests to list the most typical erroneous answers (Chasteen et al., 2012; Krijtenburg-Lewerissa et al., 2019; Mashood & Singh, 2015). Learners take the test again, with any necessary modifications made, until the developers are satisfied with the psychometric profiles of the items.

The physics education research community has established a variety of methodological techniques over the years for creating inventories in physics and science teaching. CIs have been utilised to gather data from widely diverse populations to study and compare learners' responses to each item at various skill levels and portray their behaviours using Item Response Curves (IRCs) analysis (Ishimoto et al, 2017; Richardson et al, 2021). There is a plethora of studies on how to assess and understand the outcomes of CIs in various contexts utilising contemporary methods for evaluating data using CIs. The Delphi technique, which is for identifying and quantifying expert opinions and is characterised by an iterative series of questioning and feedback, emerges frequently in the literature on creating and validating CIs in physics education research. We undertook a Delphi survey in an earlier study to modify the circular motion concept (CMC) items we used in this study (Kirya et al., 2021b).

Circular motion concepts are commonly included in the secondary school physics curriculum as learning competencies. The notions of circular motion are entrenched in the occurrences that pupils encounter daily. Previous analyses have revealed difficulties in a variety of introductory and intermediate physics courses (Mashood, 2014; Resbiantoro & Setiani, 2022; Sirait & Oktavianty, 2017). This is the precedent behind creating a circular motion concept inventory that will help assess the effectiveness of strategies for enhancing school students' conceptual comprehension. This study aimed to develop a Ugandan context-specific inventory that can recognise learners' alternative conceptions in circular motion, assess teachers' pedagogical efforts, and measure the shift in conception. It also adds to the physics education research in the Ugandan educational context while at the same time providing a model for international application.

In Ugandan schools, pupils are currently assessed mostly using traditional pen-and-paper examinations. Traditional paper-and-pencil tests provide limited ways to measure pupils' conceptual comprehension, which has been identified as a shortcoming in the Ugandan educational system (Kirya et al., 2021a). Students who come to class with misconceptions about physics or develop them while studying the subject may not have these misunderstandings addressed by teachers who provide traditional instruction and assessment. The influence of Force concept inventory (FCI) in physics education research and its success in nucleating areas of instructional innovation provided the initial push for developing a CI in the Ugandan physics context. A circular motion concept inventory (CMCI) should be a straightforward, ready-to-use tool for evaluating teaching methods and curricula. To be an effective assessment tool, it needs to establish a clear connection between concept items and their related concepts, as well as provide insights into pupil thinking by spotting misunderstandings and patterns (Kirya et al., 2021a).

Methods

Development of the Inventory

Circular Motion Key Concepts in the Ugandan Context: The National Curriculum Development Centre (NCDC) is responsible for developing a physics syllabus for Ugandan advanced secondary schools (a two-year upper secondary). Prescribed circular motion concepts are taught in the Uganda Advanced Certificate of Education (UACE) physics curriculum (NCDC, 2013). These include concepts associated with a bicycle rider, an object on a string whirled horizontally or vertically, vehicles on banked tracks, centripetal and centrifugal forces, canonical pendulum, racing vehicles around a horizontal track, rigid bodies, rotational kinetic energy, and the moment of inertia.

Development of CMC Test Items: As noted earlier, The Delphi technique was used in our earlier study, 45 items were modified from the rotational kinematics inventory and the rotating and rolling motion conceptual survey inventory (Kirya et al., 2021b). The Delphi method was chosen because it involves a consensus of opinions from experts and adheres to the methodological guidelines for developing a CI (Gerke et al., 2022; Krijtenburg-Lewerissa et al., 2019; Mashood & Singh, 2015). Nomination criteria were employed to choose physics instructors who were designated as Delphi technique experts (Clayton, 1997; Kirya et al., 2021b), in this instance university physics lecturers from the Islamic University in Uganda, Kyambogo University, and Kampala International University were nominated. The rotational kinematics inventory and the rotating and rolling motion conceptual survey inventory served as the benchmark CIs for the Delphi technique (Kirya et al., 2021a; Mashood, 2014; Rimoldini & Singh, 2005). The concept items were adapted based on the circular motion content in the Ugandan advanced physics school curriculum (NCDC, 2013). We also developed eight concept items to fill missing concepts from the benchmarked CIs, such as a bicycle rider's motion and a vehicle travelling around a circular track, using the methodological procedures for developing a CI (Kirya et al., 2021b). As a result, a total of 53 concept items were compiled for field administration.

We agreed that the number of response options should be consistent for all the 53 items. There are four possible answers in the rotational kinematics inventory. There are 5 possible answers for the rotating and rolling motion conceptual survey inventory. The adapted rotating and rolling motion conceptual survey inventory's question items had their answer alternatives reduced to four. This was in reference to the previous literature on creating CIs that suggested using a pattern of four possible answers reduces guesswork. Answer options such as “none of the above”, “all of the above”, and “others” were discarded because they were not displayed as right answers to any of the question items adapted (Planinic et al., 2019; Rimoldini & Singh, 2005).

Validation

Face Validity of the Test Items: The quality of a test to be considered as appropriate to test-takers, and test administrators is known as face validity (Sartori, 2010). The development of psychometric tests for assessment includes face validity as a key component. This offers some evidence regarding whether the test accomplishes its intended objective, which gives it credibility and scientific rigour. In this study, ten undergraduate physics students and six physics instructors, were invited to participate in the validation process before the 53 circular concept questions were administered (Chedi, 2017; Delbecq et al., 1975). The undergraduate students assessed the practicality, readability, style and formatting uniformity of the language and phrases presented. Some word choices that students found challenging included 'spinning', 'homogenous disk', 'torque', 'ice-banked curve', 'requisite', 'arbitrary', and 'skidding'. The wordings were accordingly altered.

Content Validity of the Test Items: A copy of the 53 circular concept items was given to the six nominated experts (Aslanides & Savage, 2013; Balta et al., 2022; Chasteen et al., 2012; Istiyono & Hamdi, 2020). A copy of the course content of circular motion on the UACE physics syllabus was also provided to the experts (NCDC, 2013) for reference, to be used as a blueprint copy. The experts were asked to comment on the items' relevance concerning the UACE curriculum, as well as their thoughts on its applicability, and the alternative options for each concept item representing the most common alternative conceptions for the item being examined. Eleven question items were dropped from the 53 items because they featured concepts not taught in the UACE physics syllabus, resulting in a total of 42 concept items. The experts also advised that utilising symbols to represent an alternative response is not a good way to organise options, and such items were discarded. Despite their inclusion in the physics teaching content, concept items from content areas that are not frequently taught, like a rigid uniform disk spinning through its center, were suggested for removal.

Some of the 42 concept items selected using the above-mentioned procedure were revised following the general advice of the experts. To create a successful inventory, we embraced a simple language presentation. The concept items that contained mathematical equations were also deemed to be unsuitable and discarded (Istiyono & Hamdi, 2020; Li & Singh, 2016; Mashood & Singh, 2015).

Administration of CMC Items

Participants: The 42 CMC items were answered by undergraduate students pursuing a bachelor's degree in science with a specialisation in Education, in this case Physics (BSE). Undergraduate physics students were purposively selected from four Ugandan universities: the Islamic University of Uganda, Busitema University, Kyambogo University, and Kampala International University. The participants in this study were chosen as a purposive sample since they were undergraduate students enrolled in their first semester (Mabila, 2017). Because the BSE is offered to both men and women at all Ugandan universities, the sample included both genders. The decision to take on physics undergraduate students is contingent on the satisfactory completion of a UACE physics course that covers CMCs in high school. The total number of participants was 118.

Field Test of the CMC Test Items: Physics lecturers from the four universities helped to create a testing environment for the administration of the test. Instructors at those universities were advised not to stick to the instrument's time restriction, enabling students who asked for additional time to get it. The students' performance revealed difficulties in reasoning the respective items, which we followed up with interactive student interviews. The interviews were designed for free responses to the multiple-choice items as discussed in the following section.

Students' Interactive Interview Sessions

To make sure that questions were not being misunderstood and that the correct answers were not being chosen randomly, think-aloud interviews with students were done (Balta et al., 2022; Reed & Wolfson, 2021). We conducted ten interview sessions with participating students to better understand how they thought about the circular motion concept items. This provided supplemental qualitative data to inform our interpretation of the quantitative results obtained.

When responding to open-ended versions of the same CMC items (multiple-choice options were given), the interviewers asked students to justify their answers by thinking aloud. The interviewers periodically asked follow-up questions to get a better understanding of what the students were thinking about those specific items. As a result, CMC items were thoroughly examined for their capacity to elicit actual knowledge. Overall, we discovered a lot of student issues with the CMC items.

Statistical Analysis of Data

To better analyse the data and select the concept items that performed well across the range of student abilities, our study used a mixed data analysis approach. The ITEMAN software was initially used to statistically analyse the students' scores. The programme produces descriptive statistics as output variables for data that has been programmed and run, which are then used to interpret students' overall performance. The students' performance is used to analyse CMC items' suitability based on test item characteristics. Item analysis is done to evaluate which items to retain, discard, and which need revision. The study explored item analysis in relation to the literature studies utilizing the education measurement theories of conducting item analysis to determine their appropriateness (Ding & Beichner, 2009; Sirait & Oktavianty, 2017; Suprapto et al., 2020).

The Classical Test Theory

The Classical Test Theory (CTT) is referred to as "true score theory" and its primary model states that observed test scores are made up of two components: a true score and an error score. Current psychometrics have surpassed CTT but understanding the impact of measurement quality on scientific research requires an understanding of CTT. The idea is founded on the notion that disparities in students' answers are systematic and are influenced by differences in students' abilities (DeMars, 2018; Ding & Beichner, 2009; Wong & Kanageswari, 2020). The true and error scores are not related. The idea argues that each student has a true score that can be calculated assuming no measurement errors. The context of the analysis is prescribed by the investigation of the overall test scores; i) the frequency of accurate responses indicates item difficulty. ii) the test's reliability and item-total correlation, which assesses discrimination on an item-by-item basis. iii) the frequency of responses, which is used to investigate the effectiveness of distractors. Even though these metrics are commonly used, the CTT has its drawbacks.

Item Difficulty

Utilizing the difficulty index is one method for assessing the level of difficulty in CI items. Easy, moderate, and hard are the three levels of difficulty that psychometricians have identified. The percentage of students who correctly responded to the concept item out of those who attempted it is used to calculate the difficulty indices. The mathematical formula used to determine the item difficulty index was taken into consideration in this study as

$$Difficulty\ Index = \frac{n_i}{n}$$

Where n_i is the number of students with correct responses to an item whereas$n$ is the total number of students who sat for the test (Ding & Beichner, 2009).

Reliability

The degree to which a test is repeatable and produces consistent results is known as reliability. A test's reliability is determined in a variety of methods, including several administrations of the same test several times. However, because the items of the investigation were modified from existing standardized CIs, we administered the CMC items to students in this study. We employed the standard error measurement (SEM), which shows the degree of confidence that a student's "actual" score is within a given range of values. The reliability coefficient is inversely related to SEM. So, when it comes to instrument reliability, one doesn't just think about the reliability coefficient; standard error of measurement is also an important statistical metric to consider (Tighe et al., 2010).

The ability range of students who answered the CMC item test influenced reliability. In other words, we permitted unprepared undergraduate physics students to reply to the test, believing they had been taught the CMCs and had the understanding. However, when the number of students taking the test is minimal and the sampling error affects the range of students' abilities, reliability issues arise. Because SEM is not affected by these issues, it is a superior indicator of assessment quality and is recommended for routine use (Tighe et al., 2010). Tighe et al. went on to say that when the ability range of test-takers is limited; the SEM is enough for determining measurement precision. As a result, we agree with this approach and recommend that the reliability coefficient be interpreted considering the psychometric properties and distractor functionality of CMC questions. For dependability, SEM (2.79) is preferable to the reliability coefficient.

Item Discrimination

One of the statistical analyses used by CTT and considered in this study is the discrimination index. The index is a measure of how well an item can distinguish between high and low scoring students. We aimed to select concept items that are good at discriminating students using this index. The difference between the top quartile and bottom quartile of students' percentages of correct responses to an item is used to compute the index of discrimination. We determined the discrimination index of CMC items using the students' scores and a discriminating criterion. The following formula was used to calculate the values (Ding & Beichner, 2009).

$$Discrimination\ Index = \frac{(n_u - n_l)}{\left(\frac{n}{4}\right)}$$

where n_u is the number of students with correct responses to an item in the upper quartile whereas n_l is the number of students with correct responses to an item in the lower quartile. The index has a range of -1 to 1. A satisfactory item discriminating index requires at least 0.3 rating and higher values are advised. The item with a positive discrimination index is desired (Ding & Beichner, 2009; Mashood & Singh, 2015; Wu et al., 2016). On the other hand, since the discrimination index is not always a reflection of item quality, items with discrimination indices lower than 0.30 may also be considered, provided other indices for the item are acceptable. There are several causes for an item's low discriminating power (Cohen & Swerdlik, 2009; Ding & Beichner, 2009; Wu et al., 2016).

Point - Biserial Coefficient

The item analysis often includes the point-biserial correlation coefficient (r_{pbis}) as one of the primary statistical analyses utilized by the CTT (Ding & Beichner, 2009; Mashood & Singh, 2015). The correlation assesses the relationship between students' performance on an item and their overall performance on a test (as a whole). A high value means that students are more likely to correctly answer the question if their overall test results are high. The index should be at least 0.2 and preferably higher.

Distractor Analysis

In this study, multiple-choice item distractors are used for diagnostic reasons, and when used properly, they can yield insightful diagnostic data (Gierl et al., 2017). Additionally, we assessed the effectiveness of the CMC items' distractors to determine how effective and credible they were.

The Item Response Curves

The IRT addresses some of the drawbacks of CTT. Morris et al.'s IRC is a simplified form of the IRT that allows us to have a better grasp of the quality of individual CMC items and their alternatives (Morris et al., 2006). Morris et al.'s IRCs focus on the odds of getting an item correct or incorrect based on the curves for different choices of each item. The ability level of students is related to each potential alternative option using IRCs. IRC analysis can be used for any multiple-choice evaluation beyond the typical dichotomous scheme of right/wrong, providing a strong new standard tool for generating and analysing concept items. The IRC facilitates a simultaneous comparison of both incorrect and correct answers and a deep analysis of answers at each proficiency level. "In a sense, IRCs disentangle response choices from proficiency" (Ishimoto et al., 2017). The technique also enables educators able to recognize that some correct answers are better than others. This is a very useful technique and has been used for analysing concept items for different CIs used in introductory physics, in different culture, language and educational systems (Ishimoto et al., 2017; Mashood, 2014; Morris et al., 2006; Morris et al. 2012; Rakkapao et al., 2016; Reyes & Rakkapao, 2020; Richardson et al., 2021). This study also utilised the item analysis method. Using IRCs, we plotted the percentage of pupils who selected each CMC item's response options as a function of the overall score. We were able to examine the 42 CMC items objectively and better understand the students' responses according to the IRCs. We were able to classify circular concept items that exhibited efficient, moderately efficient, and inefficient item features (Morris et al., 2006).

Analysis and Discussion of Findings

Students obtained scores ranging from 3 to 19 for a total of 42 items under investigation. There is a mean score of 30% with a standard deviation of 2.80 from this range of scores (See Table 1).

Table 1

Descriptive Statistics

Statistic	Scored Items
No of Students (N)	118
Items:	42
Mean:	12.14
SD:	2.80
Min Score:	3
Max Score:	19
Mean *p*:	0.29
Alpha:	0.00
SEM:	2.79

This is evident in the average percentage (mean p = 30%) of students who meet the item difficulty index criterion's definition of fair performance (Jalil et al., 2018). This demonstrates an inadequate understanding of the fundamental concepts underlying circular motion. The concepts were difficult because the students had deeply ingrained, nonscientific preconceptions about circular motion (Mutsvangwa, 2020). Table 2 displays the frequency with which students responded to all items on each option of the CMC items.

Table 2

Frequency of Responses to the CMC Items

CMC	Alternatives				CTT Indices			CMC	Alternatives				CTT Indices		
Item	A	B	C	D	*p*	r_{pbis}	*Dis*	Item	A	B	C	D	*p*	r_{pbis}	*Dis*
1	21	20	47	30	0.40	0.10	0.20	22	26	48	23	21	0.41	-0.10	-0.26
2	38	17	43	20	0.32	0.08	0.14	23	39	23	15	41	0.20	0.07	0.17
3	16	26	25	51	0.21	-0.14	0.08	24	2	68	43	5	0.58	0.11	-0.07
4	38	46	25	9	0.39	0.11	-0.47	25	34	45	25	14	0.29	-0.05	-0.12
5	30	7	67	14	0.12	0.02	0.01	26	9	62	10	37	0.53	0.04	0.86
6	64	12	8	34	0.54	-0.12	-0.23	27	25	10	58	25	0.21	-0.10	0.00
7	7	36	48	27	0.41	-0.02	0.27	28	14	65	25	14	0.55	-0.13	-0.07
8	14	70	17	17	0.14	-0.12	0.04	29	4	15	33	66	0.13	0.07	0.29
9	16	8	78	16	0.66	0.01	0.00	30	41	12	33	32	0.28	0.07	-0.26
10	58	11	47	2	0.49	-0.07	-0.83	31	46	44	14	14	0.12	0.24	-0.04
11	69	16	7	26	0.22	-0.15	-0.28	32	43	21	47	7	0.36	0.08	0.00
12	16	38	45	19	0.32	-0.05	-0.04	33	49	18	30	21	0.15	-0.06	-0.15
13	12	60	19	27	0.23	0.02	0.17	34	27	25	51	15	0.13	-0.05	-0.11
14	13	14	66	25	0.21	-0.21	0.07	35	15	33	37	33	0.13	0.08	0.11
15	58	8	34	18	0.15	-0.01	-0.13	36	14	11	40	53	0.34	0.13	0.50
16	41	17	11	49	0.35	-0.02	0.66	37	38	30	14	36	0.31	0.02	0.00
17	50	27	9	32	0.23	-0.15	-0.17	38	58	31	17	12	0.14	-0.07	-0.10
18	12	36	51	19	0.31	0.11	0.31	39	45	28	33	12	0.24	0.02	-0.39
19	44	32	12	30	0.10	-0.06	-0.07	40	23	38	36	21	0.32	0.01	-0.08
20	10	66	22	20	0.19	0.12	0.05	41	19	37	45	17	0.31	-0.02	0.12
21	44	40	25	9	0.21	0.11	-0.28	42	41	24	26	27	0.23	0.08	0.15

Note: The underlined corresponds to the correct alternative, p (Difficulty index), r_{pbis} (Point-biserial correlation), Dis (Discrimination index)

The findings of the study confirm widespread perceptions about CMCs (See Table 2). The relative effectiveness of concept items was measured using item difficulty (*p*). The preferred range of values for the index is [0.2: 0.75] (Jalil et al., 2018; Sirait & Oktavianty, 2017). Unsatisfactory p-values ($0.0 \leq p \leq 0.2$ and $0.75 \leq p \leq 1.00$) of CMC items indicate that the item should be disregarded. No CMC items with a high p-value ($0.75 \leq p \leq 1.00$) are identified (See Table 2). This implies that the questions were relatively difficult, which often happens with conceptual questions (Ding & Beichner, 2009; Mashood & Singh, 2015). This CMC items test found that twelve questions (29%) were difficult with $0.0 \leq p \leq 0.2$, 27 percent of items were fair ($0.0 \leq p \leq 0.2$) and 45 percent were relatively difficult ($0.26 \leq p \leq 0.74$), making them acceptable. A low *p*-value does not always imply a faulty item, because a good item can sometimes be answered erroneously by many students. This occurs when an item addresses a very strongly held alternative conception which is difficult to remedy using traditional instruction. As a result, we resolved to investigate the CMC items further, utilising the IRCs in the preceding sections to determine which CMC items should be retained.

With a score of 48%, the discrimination index rates around half of CMC items as poor. Items with a discriminating index of less than 0.20 make up 36% (n = 15) of the CMC items. Only three items are classified as marginal, being in the range of 0.20 to 0.29. Four CMC items have discrimination indices rating of good and very good in the range of 0.30 to 0.75.

Table 2 lists the r_{pbis} values for each CMC item. One item having an index higher than 0.2. Due to the successful distractors' attractiveness to high ability students, the findings of the point-biserial indices show that there was no difference between students of high and low ability levels. As a result, the outcomes were not significantly different from those of the statistical analysis of the computed discrimination indices for most of the items.

The results of this study's discrimination index analysis are different from those of many studies (Sirait & Oktavianty, 2017; Ding & Beichner, 2009) that look at psychometric indices of CTT specific to a particular field of study. It has been noted that items with inadequate and negatively discriminating questions are assessed to be of an appropriate level of difficulty. High-ability scorers incorrectly responded to these CMC items more frequently than low-ability scorers, according to a negative discriminating index, showing difficulties with the conceptual questions. The results support Planinic et al.'s (2019) claim that Croatia's high school students did not sufficiently master the material, even though it was covered. Thus, the item discrimination index does not necessarily reflect item validity, because the difficulty and discriminating indices alter whenever the CIs are administered. The effectiveness of the distractors, test instructions, the quantity of the sample tested, and the students' level of competence all have an impact on the performance of the discriminating powers of each CMC item (Cohen & Swerdlik, 2009; Wu et al., 2016).

To extract diagnostic information from CMC questions, we used the established distractor analysis technique of IRCs. The CMC item answer options are designed to provide useful diagnostic information, such as various conceptions of developmental phases related to the item. The frequency responses, graphically displayed (Histograms) in figures 1a, 1b, and 1c, were analysed for the CTT techniques (See Table 2). We next used IRCs based on IRT to evaluate CMC items and their choices beyond the typical measures of CTT indices (See Figure 1) (Morris et al., 2006).

Figure 1a

Frequency Responses of CMC Items (1 – 14)

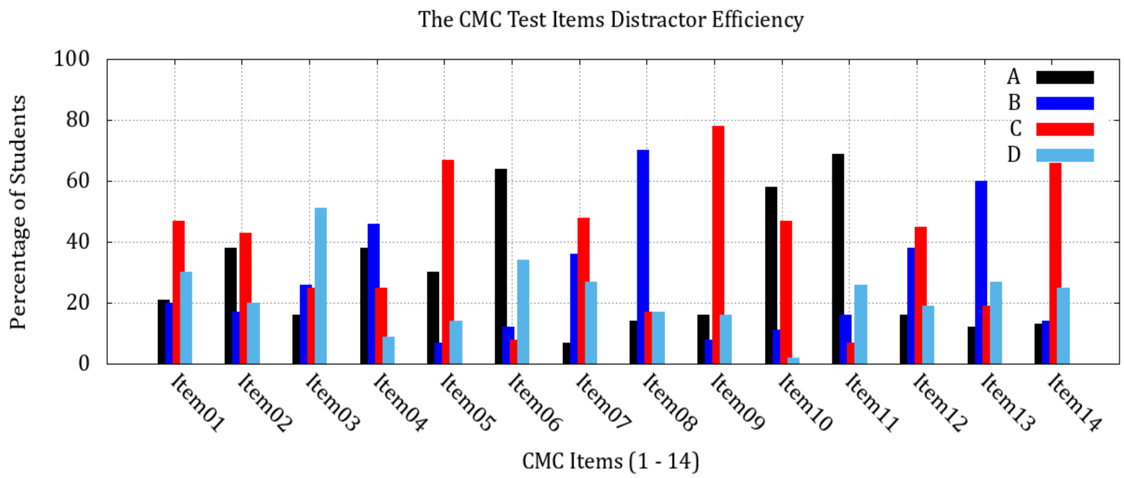


Figure 1b

Frequency Responses of CMC Items (15 – 28)

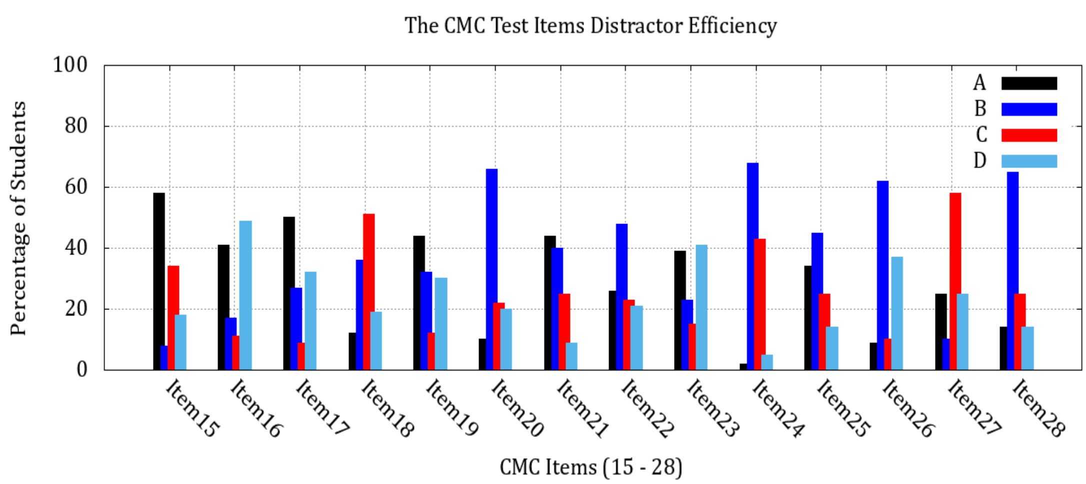
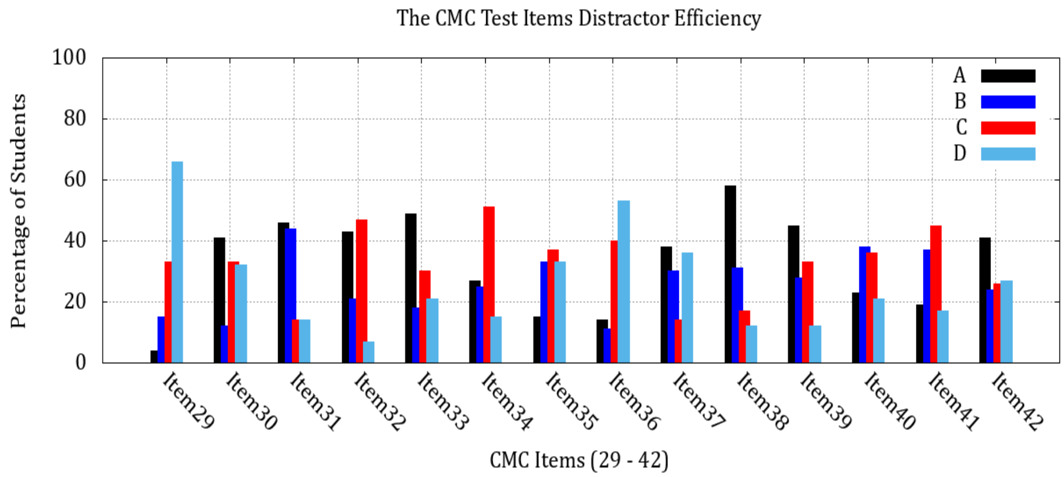


Figure 1c

Frequency Responses of CMC Items (29 - 42)



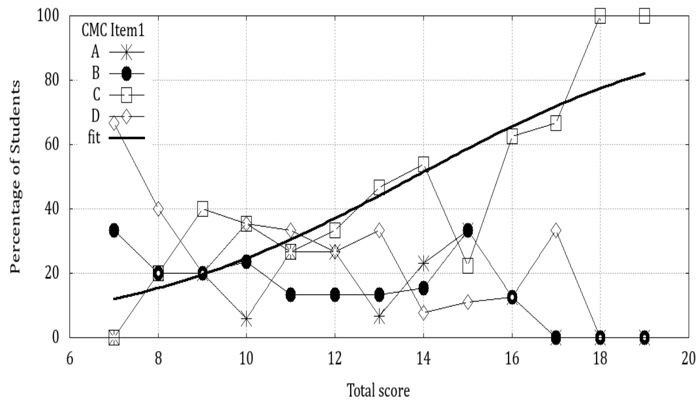
Only two of the 168 distractors were non-functional, meaning that they were selected by fewer than 5% of the respondents (CMC Items 24 & 29). One hundred sixty-six (99%) of the items were chosen by 5% or more of the students, suggesting that they are functional (Balta et al., 2022; Cizek et al., 1998). This suggests that the distractors represent widely held beliefs and are written in the students' everyday English so that they may relate to the esoteric language used in science classes (Kirya et al., 2021b).

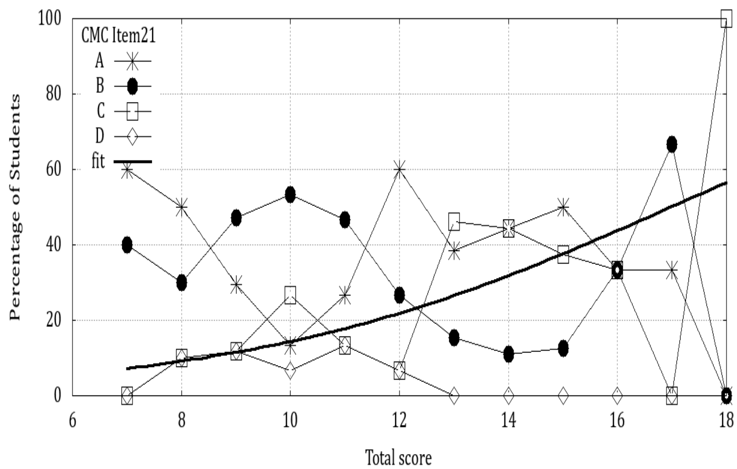
The Item Response Curves

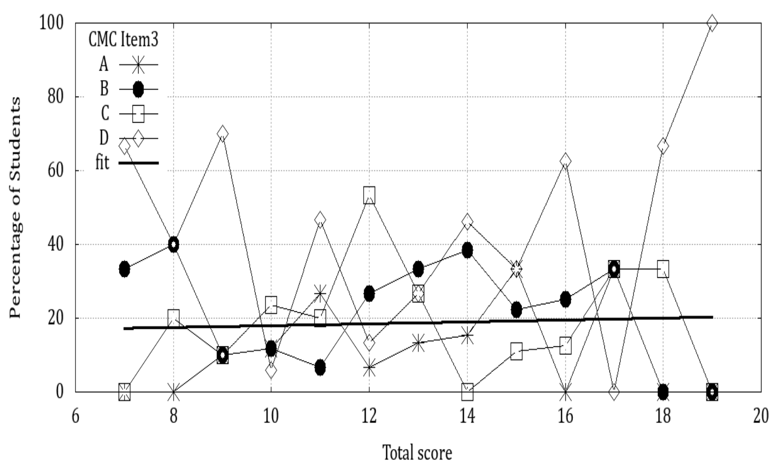
We present the IRC analysis technique for evaluating the efficacy of CMC test questions and show how it is used in this educational context. In this work, a few IRCs for CMC items are presented and discussed (1, 21 & 3). When comparing the final score to each response option, we utilized the proportion of students who selected each alternative (See Figure 2).

Figure 2

IRCs (CMC.Item 1, 21 & 3)







We utilised the logistic response function as a model for a goodness-of-fit of the correct responses to make items easier to identify and analyze (Kucharavy & De Guio, 2015). As a first illustration, using CMC.Item1, the sigmoid curve shows that low-ability students have significant misconceptions because their non-correct responses are higher than those of high-ability students. The curve matches the shape of efficient questions that are discriminating, as described by Morris et al. (2006). The other items 2, 16, 18, 24, 26, 32, 39, and 40 also have this characteristic.

Using the moderately challenging CMC.Item21 as a second example, 38 (32.2%) students chose the correct score (Alternative B). Following Morris et al. (2006), we refer to item 12 as a moderately efficient question due to this finding. We observed the following items having the same characteristic among the CMC items: 6, 9, 7, 12, 13, 20, 23, 28, 30, 31, 36, and 37. The slope of the linear form of the sigmoid curve indicated that there was a minimal separation between low and high-ability students in the correct score within the items. The distractors are more revealing about student abilities for these items, as evidenced by their IRCs.

As the third example, we analysed the distractor IRCs with the appropriate fit to the correct score in our analysis of CMC.Item3 (Alternative C). There is no difference in ability between students with low and high ability, according to the item's contrast of the IRCs for the fitted correct score to the distractors (A, B, & D). This is shown by the almost linear fitted correct score across the whole range [6:19]. Such items are referred to be inefficient items with properties of this character by Morris et al. (2006). Such items are difficult for students to understand and do not provide useful information during analysis. Items 4, 5, 8, 10, 11, 14, 15, 17, 19, 22, 25, 27, 29, 33, 34, 35, 38, 41, and 42 are those exhibiting this character. Follow up on these questions with a group of students (N=10) occurred during interview interaction sessions. As a result, the inefficient items' IRCs revealed that fitting correct scores were not discriminatory.

Items that are regarded to be efficient and moderately efficient are used to compile a CMCI. Items 1, 2, 6, 7, 9, 12, 13, 16, 18, 20, 21, 23, 24, 26, 28, 30, 31, 32, 36, 37, 39 and 40 meet this criterion. The correct responses to the CMC items are related to the students' competence levels, as shown in their IRCs. The IRC technique is recommended since it identifies item test distractors that are not visible using traditional analytical methods (Morris et al., 2006) at various levels of competence of test-takers. As a result, the compiled CMC items are appropriate for evaluating students' conceptual understanding because they are designed to target different cognitive skill levels and can be utilized to spot misconceptions among students and areas that require specialized instruction.

Conclusion

This study, which is guided by the psychometric measures of CTT and IRC, documents the empirical support for developing the CMCI. SEM (2.79) is preferred over the reliability coefficient in terms of dependability. The 42 CMC items in this educational setting were evaluated using the psychometric and IRC analyses, to reveal well-performing items. The CTT analysis determined that 26.2% of the items had a fair difficulty level and 45.2% had a moderate difficulty level. The IRC analysis is utilized to develop a 22-item CMCI. Physics instructors from Uganda and the East African region can utilize this CMCI to evaluate students' understanding of circular motion concepts. This CMCI can be used to assess students' understanding of circular motion concepts and compare results from pre-test and post-test across curricula to determine students' gains. Further studies of administering the CMCI for determining misconceptions on a large population by physics instructors from Uganda and the East African region can also be carried out.

Acknowledgements

We appreciate the physics educators and undergraduate physics students for their valuable time and contribution to this study. The financial assistance from ACEITLMS is greatly acknowledged.

References

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, *33*(9), 1289-1312. https://www.tandfonline.com/doi/abs/10.1080/09500693.2010.512369

Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). *Introduction to research in education*. Cengage. http://repository.unmas.ac.id/medias/journal/EBK-00124.pdf

Aslanides, J. S., & Savage, C. M. (2013). Relativity concept inventory: Development, analysis, and results. *Physical Review Special Topics- Physics Education Research*, *9*(1), 010118. https://link.aps.org/doi/10.1103/PhysRevSTPER.9.010118

Balta, N., Dzhapashov, N., Salibasic, D., & Mesic, V. (2022). Development of the High School Wave Optics Test. *Journal of Turkish Science Education,* *19*(1), 306-331. https://files.eric.ed.gov/fulltext/EJ1343220.pdf

Chasteen, S. V., Pepper, R. E., Caballero, M. D., Pollock, S. J., & Perkins, K. K. (2012). Colorado upper-division electrostatics diagnostic: A conceptual assessment for the junior level. *Physical Review Special Topics-Physics Education Research*, *8*(2), 020108. https://link.aps.org/doi/10.1103/PhysRevSTPER.8.020108

Chedi, J. M. (2017). A preliminary review on needs analysis and Delphi technique: Effective tools for data collection. *Journal of Asian Vocational Education and Training*, *10*, 44-52. http://www.javet.net/wp-content/uploads/2019/11/010-04-2017_Jamilu-Mustapha-Chedi_A Preliminary-Review-on-Needs-Analysis-and-Delphi-Technique.pdf

Cizek, G. J., Robinson, K. L., & O'Day, D. M. (1998). Nonfunctioning options: A closer look. *Educational and Psychological Measurement, 58*(4), 605-611. https://doi.org/10.1177/0013164498058004004

Clayton, M. J. (1997). Delphi: a technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology*, *17*(4), 373-386. https://www.tandfonline.com/doi/abs/10.1080/0144341970170401

Cohen, R., & Swerdlik, M. (2009). *Psychological Testing and Assessment: An Introduction to Tests and Measurement (7th ed.).* New York: McGraw Hill.

Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H. (1975). *Group techniques for program planning: A guide to nominal group and Delphi processes*. Scott, Foresman, https://journals.sagepub.com/doi/abs/10.1177/105960117600100220

DeMars, C. E. (2018). Classical test theory and item response theory. *The Wiley handbook of psychometric testing*: *A multidisciplinary reference on a survey, scale, and test development*, 49-73. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118489772.ch2

Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics - Physics Education Research*, *5*(2), 020103. https://link.aps.org/doi/10.1103/PhysRevSTPER.5.020103

Gerke, F., Müller, R., Bitzenbauer, P., Ubben, M., & Weber, K. A. (2022). Requirements for future quantum workforce – a Delphi study. *Journal of Physics: Conference Series 2297*(1), 012017. https://iopscience.iop.org/volume/1742-6596/2297

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*. *87*(6), 1082–1116. https://doi.org/10.3102/0034654317726529

Ishimoto, M., Davenport, G., & Wittmann, M. C. (2017). Use of item response curves of the Force and Motion Conceptual Evaluation to compare Japanese and American students' views on force and motion. *Physical Review Physics Education Research*, *13*(2), 020135. https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.13.020135

Istiyono, E., & Hamdi, S. (2020). Measuring Creative Thinking Skills of Senior High School Male and Female Students in Physics (CTSP) Using the IRT-based PhysTCreTS. *Journal of Turkish Science Education*, *17*(4), 578-590. https://files.eric.ed.gov/fulltext/EJ1283951.pdf

Jalil, S., Ali, M. S., & Haris, A. (2018). Development and validation of science process skills instrument in physics. *Journal of Physics: Conference Series*, *1028* (1), 012203. https://iopscience.iop.org/volume/1742-6596/1028

Kirya, K. R., Mashood, K. K., & Yadav, L. L. (2021a). Review of research in student conception studies and concept inventories- Exploring PER threads relevant to Ugandan context. *African Journal of Educational Studies in Mathematics and Sciences*. *17*(1), 37-60. https://www.ajol.info/index.php/ajesms/article/view/219548

Kirya, K. R., Mashood, K. K., & Yadav, L. L. (2021b). A methodological analysis for the development of a circular-motion concept inventory in a Ugandan context by using the Delphi technique. *International Journal of Learning, Teaching and Educational Research*. *20*(10), 61-82. https://doi.org/10.26803/ijlter.20.10.4

Krijtenburg-Lewerissa, K., Pol, H. J., Brinkman, A., & Van Joolingen, W. R. (2019). Key topics for quantum mechanics at secondary schools: a Delphi study into expert opinions. *International Journal of Science Education*, *41*(3), 349-366. https://www.tandfonline.com/doi/abs/10.1080/09500693.2018.1550273

Kucharavy, D., & De Guio, R. (2015). Application of logistic growth curve. *Procedia engineering*, *131*, 280-290. https://doi.org/10.1016/j.proeng.2015.12.390

Laverty, J. T., & Caballero, M. D. (2018). Analysis of the most common concept inventories in physics: What are we assessing? *Physical Review Physics Education Research*, *14*(1), 010123. https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.14.010123

Li, J., & Singh, C. (2016). Developing and validating a conceptual survey to assess introductory physics students' understanding of magnetism. *European Journal of Physics*, *38*(2), 25. https://doi.org/10.1088/1361-6404/38/2/025702

Mabila, T. E. (2017). Postgraduate students' understanding of mixed methods research design at the proposal stage. *South African Journal of Higher Education*, *31*(5), 136-153. https://doi.org/10.20853/31-5-1498

Mashood, K. K., (2014). Development and evaluation of a concept inventory in rotational kinematics. Unpublished Doctoral dissertation, Tata Institute of Fundamental Research, Mumbai. Retrieved January 18, 2020, from http://www.hbcse.tifr.res.in/research-development/ph.d.-theses/thesis mashoodkk.pdf

Mashood, K. K., & Singh, V. A. (2015). Rotational kinematics of a rigid body about a fixed axis: development and analysis of an inventory. *European Journal of Physics*, *36*(4), 045020. https://iopscience.iop.org/article/10.1088/0143-0807/36/4/045020/meta

Morris, G. A., Branum-Martin, L., Harshman, N., Baker, S. D., Mazur, E., Dutta, S., Mzoughi, T., & McCauley, V. (2006). Testing the test: item response curves and test quality. *American Journal of Physics. 74*, 449–53. https://doi.org/10.1119/1.2174053

Morris, G. A., Harshman, N., Branum-Martin, L., Mazur, E., Mzoughi, T., & Baker, S. D. (2012). An item response curves analysis of the Force Concept Inventory. *American Journal of Physics*, *80*(9), 825-831. http://dx.doi.org/10.1119/1.4731618

Mutsvangwa, A. (2020). A study of student teachers' misconceptions on uniform circular motion. *Journal of Physics: Conference Series*, *1512* (1), 012029. https://doi.org/10.1088/1742-6596/1512/1/012029

National Curriculum Development Centre (NCDC), Uganda (2013). Uganda Advanced Certificate of Education, Teaching Syllabi for Physics. Retrieved January 20, 2020, from https://www.ncdc.go.ug/

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 020111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Rakkapao, S., Prasitpong, S., & Arayathanitkul, K. (2016). Analysis test of understanding of vectors with the three-parameter logistic model of item response theory and item response curves technique. *Physical Review Physics Education Research*, *12*(2), 020135. https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.12.020135

Reed, C. R., & Wolfson, A. J. (2021). Concept inventories as a complement to learning progressions. *CBE—Life Sciences Education*, 20(2), es4. https://www.lifescied.org/doi/abs/10.1187/cbe.20-09-0208

Resbiantoro, G., & Setiani, R. (2022). A review of misconception in physics: the diagnosis, causes, and remediation. *Journal of Turkish Science Education*, *19*(2). https://www.tused.org/index.php/tused/article/download/924/773

Reyes, M. G., & Rakkapao, S. (2020). Item response curve analysis of Likert scale on learning attitudes towards physics. *European Journal of Physics*, *41*, 045703. https://doi.org/10.1088/1361-6404/ab805c

Richardson, C. J., Smith, T. I., & Walter, P. J. (2021). Replicating analyses of item response curves using data from the Force and Motion Conceptual Evaluation. *Physical Review Physics Education Research, 17*(2), 020127. https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.17.020127

Rimoldini, L. G., & Singh, C. (2005). Student understanding of rotational and rolling motion concepts. *Physical Review Special Topics - Physics Education Research*, *1* (010102), 1- 9. https://doi.org/10.1103/PhysRevSTPER.1.010102

Sartori, R. (2010). Face validity in personality tests: psychometric instruments and projective techniques in comparison. *Quality & Quantity*, *44*(4), 749-759. https://doi.org/10.1007/s11135-009-9224-0

Sirait, J. H., & Oktavianty, E. (2017). Analysis of pre-service physics teachers' understanding of vectors and forces. *Journal of Turkish Science Education, 14*(2), 82-95. https://www.academia.edu/34130543/Analysis_of_Pre_Service_Physics_Teachers_Understanding_of_Vectors_and_Forces

Suprapto, E., Sumiharsono, R., & Ramadhan, S. (2020). The Analysis of Instrument Quality to Measure the Students' Higher Order Thinking Skill in Physics Learning. *Journal of Turkish Science Education*, *17*(4), 520-527. https://files.eric.ed.gov/fulltext/EJ1284007.pdf

Tighe, J., McManus, I. C., Dewhurst, N. G., Chis, L., & Mucklow, J. (2010). The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP (UK) examinations. *BMC Medical Education*, *10*(1), 1-9. https://doi.org/10.1186/1472-6920-10-40

Wong, V., & Kanageswari, S. S. S. (2020). The role of classical test theory to determine the quality of classroom teaching test items. *Pedagogia: Jurnal Pendidikan*, *9*(1). https://doi.org/10.21070/pedagogia.v9i1.123

Wu, M., Tam, H. P., & Jen, T. H. (2016). Educational measurement for applied researchers. Theory into practice, 136. https://doi.org/10.1007/978-981-10-3302-5