Investigating the Applicability of the Newtonian Gravity Concept Inventory to First Year College Students

Hisham N. Bani-Salameh[1,*], Ahmad Subahi[2], Sami H. Alharbi[1], Nouf Adegaither[1], Mufeed Awawdeh[2], Dalal Alamoudi[2] & Reem Alamoudi[2]

[1]College of Science and Health Professions, King Saud Bin Abdulaziz University for Health sciences, King Abdullah International Medical Research Center, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia

[2]College of Science and Health Professions, King Saud Bin Abdulaziz University for Health sciences, King Abdullah International Medical Research Center, Ministry of National Guard Health Affairs, Jeddah, Saudi Arabia

[*]Corresponding author: Mail Code (3124), PO Box 3660, Riyadh 11481, Saudi Arabia. E-mail: salamehh@ksau-hs.edu.sa

Received: October 27, 2021 Accepted: December 27, 2021 Published: February 15, 2022

doi:10.5296/ije.v14i1.19122 URL: https://doi.org/10.5296/ije.v14i1.19122

Abstract

We report on a study of our students' understanding of gravity using the Newtonian Gravity Concept Inventory (NGCI). This article is supposed to serve two purposes: as a proof of applicability of the test to first year medical students outside the United States and as a general survey of our students' understanding of gravity. The motivation for this work came initially from students' misconceptions of gravity and related topics noticed in classrooms. NGCI has 26 multiple-choice questions probing students' knowledge of gravity in four different domains: Directionality, Gravity as a force, Independence from other forces and threshold. Results confirmed weak overall performance of the 684 students participated in this study (average score of only 38.48%) with misconceptions related to gravity in all four domains. We were able to prove the applicability of the NGCI to our students through calculations of Classical Test Theory statistics and Cronbach's alpha. We got a Cronbach's alpha of 0.68, average difficulty index of 0.62 and average discriminatory index of 0.31 with three questions reported as too difficult based on the acceptable range. The test was given to students twice; before and after any gravity-related instructions in class (pre- and post-test). Post-test results will not be discussed here and left for future articles. In this article, we present the pre-test results and compare it to the original work by the authors of the test on astronomy and physics students.

Keywords: Newtonian Gravity Concept Inventory, gravity, physics education research, medical students, first year college students

1. Introduction

Gravity plays an important role in our everyday lives (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Williamson & Prather & Willoughby, 2016; Smith & Treagust, 1988), from the simple act of dropping something from one's hand to more complicated rockets launching to outer space. This is one reason for us to research students' understanding of gravity, but the more important motivation for this work came from classrooms. After teaching this class for years, we noticed many gravity-related misconceptions floating around based on interactions with students and test's records.

Just like any other concept, people will always have some firm well established wrong ideas and understandings that come from years of learning in and outside schools (Halloun & Hestenes, 1985a). Unless intercepted at one point, these misconceptions tend to accumulate over time and get deeper and more profound and consequently more difficult to correct. College students are no exception, and unless we do something about this, students' performance in college will be affected and their prior college misconceptions stay with them. Knowing what ideas about different concepts students bring with them to classrooms is crucial to instructors all around the world to improve the teaching and learning process. This will make it much easier to change or just modify classroom's instructions and ways of teaching. If done correctly, this will encourage students to think in the right direction and realize their own misconceptions and hopefully replace them with the correct ones.

To acquire this knowledge of students' misconceptions in different areas, one will always need a reliable assessments tool. When dealing with large numbers of students, the tool of choice is a concept inventory (CI) with multiple choice questions (Williamson, 2013; Williamson et al., 2016; Bani-Salameh, 2017a, 2017b, 2017c; Bani-Salameh & Nuseirat &Alkofahi, 2017a, 2017b; Ding & Beichner, 209). Anything with other than multiple choice questions will be too difficult and impractical to implement when dealing with large numbers. The answer choices in a good concept inventory are chosen very carefully based on research to form the so called "distracters" (Williamson, 2013; Bani-Salameh, 2017a; Haladyna & Downing & Rodriguez, 2002). Each distracter in every question correspond to a certain known misconception and therefore students will be tempted to choose it if they carry that misconception. This is the beauty of using CI's; you get useful information not only from correct answers, but from wrong ones as well.

With concept inventories, researchers and instructors also have the power to evaluate the effectiveness of the whole educational process they're conducting (teaching methods, curricula, text books etc.) (Williamson, 2013; Williamson et al., 2016; Bani-Salameh, 2017a, 2017b, 2017c; Bani-Salameh et al., 2017a, 2017b; Ding & Beichner, 2009). This is possible by administering the test to students twice, before any classroom's topic-related instructions (pre-test) and once afterward (post-test). Typically, student's performance in both tests will be different (hopefully better in the post test) and therefore one can calculate the gain. The gain in students' performance is a direct measure (among other things) of the effectiveness of the teaching instructions.

2. Methodology

This research project is part of a bigger project we started few years back to assess the educational process at our university in its two dimensions: teaching and learning (Bani-Salameh, 2017a, 2017b, 2017c; Bani-Salameh et al., 2017a, 2017b). Here we chose the Newtonian Gravity Concept Inventory (NGCI) as our assessment tool (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Williamson et al., 2016). The test was given to a total of 684 students in 9 different section located in two different campuses of our university. Each one of these students took the test twice as a pre- and as a post-test. The NGCI consists of 26 multiple-choice questions that probe students' understanding of gravity in four different domains: Directionality, Gravity as a force, Independence from other forces and threshold.

The directionality domain as the title suggests, probes students' ability to determine the direction of the gravitational force. Questions in this domain ask about the direction in different situations like an object resting on the surface of a large body or one with multiple objects involved. In the "Gravity as a force" domain, questions investigate students' ability to determine the magnitude of the gravitational force and its dependence on mass and distance. The idea of the "Independence from other forces" domain is to check for the known misconception of mixing the force of gravity with other forces related to earth magnetism, rotation and air pressure (Hestenes, Wells & Swackhamer, 1992; Gunstone & White, 1981; Piburn, 1998; Feeley, 2007; Asghar & Libarkin, 2010). The last domain (Threshold) is for testing students' understanding of the simple idea of the gravitational force being universal and its existence regardless of how small or big the distances or the masses are. It also tests for the misconception of the existence of a limit for the force of gravity where it will suddenly stop, some students think such a limit do exist at the edge of the atmosphere (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Williamson et al., 2016).

All of our students are medical students in their first year of college with gravity knowledge limited to what they acquired over the years from schools and from real life experiences. The course we teach them is a general physics course that's required as part of their school program. We teach them the concept of gravity in one or two lectures and we chose to test their understanding of this concept because its importance goes far beyond the two lectures. This concept is needed and applied over and over again throughout the semester in treatments of many other concepts like Newton's laws, conservation laws, the treatment of weight as a force and others.

NGCI was developed initially for introductory college astronomy students (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Haladyna et al., 2002; Benson & Clark, 1982) but the authors later proved that it's just as reliable and robust for introductory college algebra-based physics students (Williamson et al., 2016). They discussed their results of the application of the Classical Test Theory (CTT) and argued that the test's items still have appropriate difficulty and discriminating capabilities for physics students. In this work we are adding to their evidence of the applicability of the NGCI to a wide range of college students by giving the test to medical students outside the United States. We present CCT statistics for our students and compare it to the results from the original work both for astronomy and physics students.

Based on the fact that physics students are different from astronomy students in many ways and that the CTT quantitative evaluation is highly sample dependents (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Williamson et al., 2016), authors of the NGCI decided to go through test validation process again for physics students. Comparing our students with student in the United States in general and with students in the original work in particular, one realizes quickly how different they are: science and math background, academic institutions attended, motivations, interests, Way of life among other things. According to the developers (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Williamson et al., 2016), these differences will affect students' correct answer choice for different questions in a complicated way that's hard to measure. Therefore, we decided to check the applicability of the NGCI to our students and go through the validation process again by calculating the Cronbach's Alpha, Difficulty and Discrimination power for all questions in the test as discussed in detail in the next section.

3. Reliability and Validity of the NGCI for Our Students

As discussed in the methodology section, we used the NGCI to test our students' understanding of gravity and we decided to check its reliability as an assessment tool for them first (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Williamson et al., 2016). To make it easier to compare with the original work, we did the same statistical analysis: simple standard deviation and mean calculations and the use of Classical Test Theory (CTT) quantitative evaluation along with the calculation of Cronbach's alpha (α) for validation and reliability measurements. CTT is one of the easiest to implement and most commonly used method of statistical analysis for evaluation of multiple-choice tests (Williamson, 2013; Williamson et al., 2016; Ding & Beichner, 2009; Benson & Clark, 1982; Schlingman et al., 2012; Wallace & Bailey, 2010).

As for the Cronbach's Alpha (α), people usually report it as a measure of reliability. Possible values for α ranges between 0 and 1 with the zero meaning that all test's items are completely independent and higher values indicating the variance of total test scores is large compared to the variance within each item. For any assessment tool, higher values of α mean that the test is able to measure differences among students rather than just evaluating how different test's items are; this is what one wants from a good reliable test. There are different opinions about the acceptable values of alpha with no clear agreement on the labels used to describe them (Wallace & Bailey, 2010; George & Mallery, 2009; Taber, 2018). Taber (2018) documented most of the up-to-date used labels with their ranges of values for the α. Based on Taber's records; α values between 0.6 and 0.7 were described as acceptable, satisfactory and sufficient by different scholars (Taber, 2018). Taber also suggested that lower values of alpha for a certain test shouldn't always mean an unsatisfactory test. Indeed, some scholars used data collected with instruments with α of only 0.5 (Griethuijsen et al., 2014) arguing that α can be increased by simply increasing the number of items in the instrument. We calculated Cronbach's alpha based on the equation:

$$\alpha = \frac{k}{k-1}(1-\left(\frac{\sum_{i=1}^{k}\sigma_{y_i}^2}{\sigma_x^2}\right))$$

Where k is the number of students participating in the test, σ_x^2 is the variance of total scores, $\sigma_{y_i}^2$ is the variance of individual component on the test and $x = y_1 + y_2 + \cdots + y_k$ with *x* being the total test score and *y* being score on individual test's item.

There are different ways to calculate the item difficulty and discrimination. We followed the same definitions used by the test's developer to make it easier and more useful to compare. The difficulty index (*D*) for a certain item was calculated as the ratio of the number of students who got that item incorrect to the total number of students meaning that higher values of *D* reflect a difficult question. *D* for most of the items should have values between 0.2 and 0.8 (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Williamson et al., 2016; Schlingman et al., 2012) in order for the test to be with appropriate difficulty for students, not too hard and not too easy.

Lastly, item discrimination is a measure of how well a certain item is being able to differentiate students who have a certain misconception from those who don't. To find the discrimination index for a certain item, we used the point-biserial calculations to find correlations between students' performance on that item to their overall performance on the NGCI (Williamson et al., 2016; Wallace & Bailey, 2010; Lord & Novick, 1968). The point-biserial index is calculated based on the equation:

$$r_{pb} = \frac{s_1 - s_0}{SD}\sqrt{\frac{n_1 n_0}{n^2}}$$

Where for a certain item on the test, s_1 and s_0 are the average scores of students answering that item correctly and incorrectly, respectively. n_1 and n_0 are the number of students answering that item correctly and incorrectly, respectively. *SD* is the standard deviation and *n* is the total number of students. Possible values for r_{pb} range from *-1* to *1* with highly discriminating items having value ≥ 0.3, items with values between 0.2 and 0.29 are considered as marginal and below 0.2 as poor (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Williamson et al., 2016; Schlingman et al., 2012; Wallace & Bailey, 2010).

4. Results

Our calculation of the Cronbach's alpha gave a value of *0.68*, this is lower than both astronomy and physics groups in the original work (astronomy $\alpha = 0.79$ and physics $\alpha = 0.82$) (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Williamson et al., 2016). Cronbach's alpha is sample dependent (Williamson, 2013; Williamson et al., 2016; Wallace & Bailey, 2010; Lord & Novick, 1968; Hambleton & Jones, 1993) and therefore, it's of no surprise that our alpha is different. Based on the discussion we presented in the previous section, our judgment is that the test is still internally consistent and reliable for our student and therefore, it's still sensitive to differences among students rather than differences across test questions.

The average score of the 684 students who took the test was 38.5 % with a standard deviation of 14.4. Averages of students from the original work were higher: astronomy students (pre-test) 43.7 % with *SD* = 19.0 (Williamson & Willoughby, 2012, 2013; Williamson, 2013) and physics students (pre-test) 58.0 % with SD 18.3 (Williamson et al., 2016). Our students' performance seems to be closer to that of the astronomy students where it's supposed to be closer to the physics students' performance. Physics students in the original work were in algebra based introductory physics class and some of them are medical students just like ours. We have to keep in mind that our students are different than the students in the original study in so many ways: background, life experiences, way of life, institutions attended which means different curricula, different instructors and different way of teaching. All of these factors among other things will always affect students' performance and can explain the lower scores (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Williamson et al., 2016).

Table 1. CTT Statistics for Each Question in the NGCI Including Difficulty and Discrimination Indices

item	***D***	r_{bs}	***item***	***D***	r_{bs}	***item***	***D***	r_{bs}
1	0.45	0.18	10	0.90	0.24	18	0.58	0.38
2	0.83	0.36	11	0.49	0.33	19	0.60	0.40
3	0.71	0.29	12	0.54	0.34	20	0.56	0.46
4	0.76	0.31	13	0.58	0.33	21	0.72	0.40
5	0.29	0.29	14	0.62	0.34	22	0.58	0.40
6	0.57	0.25	15	0.73	0.31	23	0.73	0.41
7	0.54	0.15	16	0.70	0.16	24	0.62	0.30
8	0.75	0.44	17	0.55	0.33	25	0.85	0.07
9	0.27	0.25				26	0.46	0.36

Table 1 shows CTT statistics (both difficulty and discriminatory indices) for all questions in the test. In the original work (Williamson et al., 2016), authors reported one question to be too difficult (*Q25: D = 0.82,* r_{pb} *= 0.34*) and one question to be too easy (*Q5: D = 0.13,* r_{pb} *= 0.39*) for physics students. *Q10* for these students was very close to be considered as too difficult with *D = 0.77.* As for the astronomy students (Williamson & Willoughby, 2012, 2013; Williamson, 2013), two questions were reported as too difficult (*Q25: D = 0.89,* r_{pb} *= 0.12* and *Q10: D = 0.84,* r_{pb} *= 0.25*) and none as too easy. It seems to be the same general story with our students with results closer to the astronomy group than the physics. All questions have appropriate difficulty ((Williamson et al., 2016; Schlingman et al., 2012; Wallace & Bailey, 2010; Lord & Novick, 1968) except for: (*Q25: D = 0.85,* r_{pb} *= 0.07, Q10: D = 0.90,* r_{pb} *= 0.24* and *Q2: D = 0.83,* r_{pb} *= 0.36*) reported as too difficult with none reported as too easy. We also found some questions with low r_{pb} with question 25 being the lowest (0.07) meaning it's not a good discriminator for students' abilities. Results like these with diverse groups of students finding same questions to be difficult and same questions to be easy reflect same way of thinking and same misconceptions regarding gravity

Based on calculations of Cronbach's alpha for reliability and based on point-biserial correlations for difficulty and discrimination indices for all questions on the NGCI for our students, and based on the broad range of test scores, we conclude that the NGCI is a reliable and robust instrument to assess students' understanding of gravity. This conclusion just adds to its already established reliability and farther proves its applicability to a wide range of students.

Table 2 summarizes our results and compares it to results from original work (Williamson & Willoughby, 2012, 2013; Williamson, 2013; Williamson et al., 2016). One notice quickly the weaker performance of our students compared to both groups in the original work (astronomy and physics students). Our students total score average is *34 %* less than that of physics students and *14 %* less than that of astronomy students. This relates directly to (and somewhat is explained by) the average difficulty indices shown in table 1.

Average difficulty values for all questions as measured for our students was *48 %* and *13 %* higher than that for physics and astronomy students respectively. This is clear evidence that the test was more difficult for our students than the others indicating a weaker background and more settled misconceptions in their minds. Lastly, all averages of discrimination indices in table 2 seem to be close to each other with that for our students being the lowest. This is telling us that the test served as a very good discriminator of high ability students from low ability ones

Table 2. Comparison of the Mean, Standard Deviation, Cronbach's Alpha, Average Difficulty Index and Average Discrimination Index for Our Students with Students from the Original Work

	Mean %	**SD %**	**α**	**Average D**	**average r_{pb}**
Our students	38.48	14.39	0.68	0.62	0.31
Astronomy students Original work	43.71	19.01	0.79	0.55	0.39
Physics students Original work	58.00	18.31	0.82	0.42	0.45

5. Discussion

In this section we present more details about the overall performance of our students represented by normalized frequencies for both students' scores and individual test's items. Figure 1 shows the normalized frequencies for each question in the test calculated as a ratio of the number of students who answered the question correctly to the total number of students. This graph is basically a representation of questions' difficulties as one can clearly see the questions that students had trouble with are (Q2, Q10 and Q25) just as reported in the description of the difficulty indices. It's also worth noticing that students from original work had trouble with the same questions. Figure 1clearly indicates that students struggled with all items in the NGCI except for questions: 5, 9 and maybe 1 and 26. On 22 out of the 26 total questions, less than 50 % of all students were able to answer correctly with that percentage reaching down to 10 % for some questions.

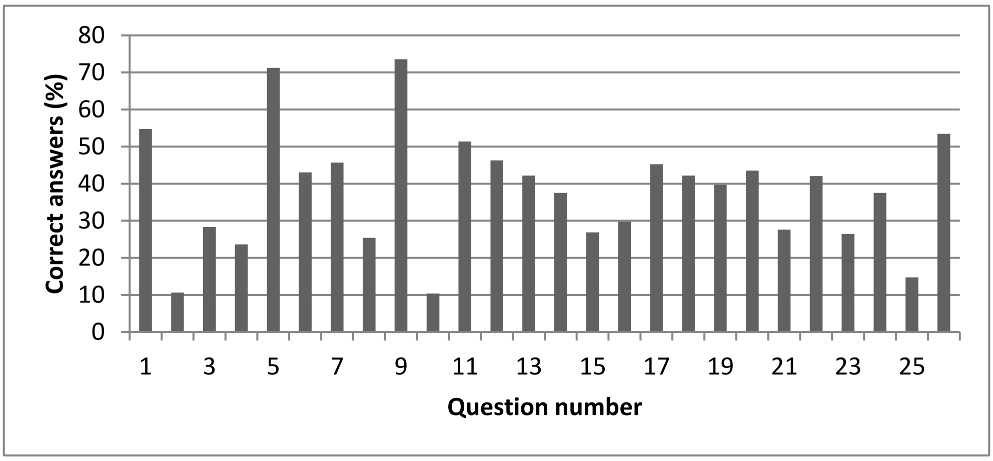

Figure 1. Correct Response Normalized Frequency for Each Item on the NGCI

In figure 2, we see normalized scores frequencies calculated as the ratio of the number of students with a certain number of correct answers to the total number of students. The general distribution is very close to normal bell-shaped with a little tail to the right of the center. This is a supporting evidence of weak overall performance of our students on the test that agrees well with information from table 1 and figure 2. It's also an indication of a good test that was able to discriminate high ability students from low ability ones.

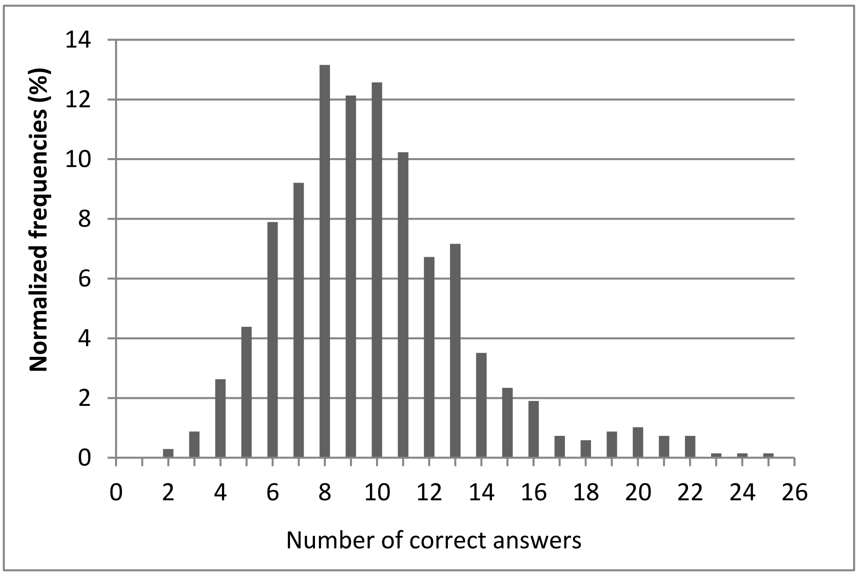

Figure 2. Students' Scores Normalized Frequencies on the NGCI

6. Conclusions

The purpose of this work was to serve as an initial investigation of students' understanding of gravity as part of a bigger project at our university. The other purpose is to test the applicability of the NGCI to our students. The results gave us very good information regarding students understanding of gravity and alarmed us of the weak performance of students. In this report, we presented only the pre-test data just to give the idea of students' overall performance on the NGCI. In our future reports, we'll look deeper into misconceptions held by students and their level of persistence by comparing students' performance in the post-test to the pre-test and calculating the gain. As for the results presented in this report, we noticed general overall similarities of our results with results from the original work both for astronomy and physics students with ours being closer in performance to that of astronomy students. This reflects similar way of thinking and same misconceptions held regardless of big differences among them.

As for the second purpose of this report, our calculation of the Cronbach's alpha was 0.68 which is lower than reported in the original work. We also reported the difficulty indices and the discrimination indices for each item on the test with average difficulty index of 0.62 that prove again the weak performance of our students. Based on the detailed discussion we presented in the reliability section we conclude that the test is still a reliable and robust instrument for testing our students' understanding of gravity. This report should serve as the first report on applying the NGCI outside the United States and add to its applicability to a diverse population regardless of their background, country of origin, motivations, interests in life and any other differences.

In figure 1 and figure 2 we reported normalized frequencies of students' scores and individual test items. These two graphs along with the low total score average of only 38.48 % is an indication of weak overall performance of students. In figure 2 we see a positively skewed close-to-normal distribution which is an indication of a good test. In figure 1 it's clear that a large percentage of students were struggling with most of the questions and got them wrong at the end. Our initial conclusion from these results is that our students carry a large number of misconceptions related to the gravity concept. In order to be able to help students replace these wrong ideas, we need to determine them first and then measure the effectiveness of our classroom instructions on them. This will be done in the near future and will be reported in the following article.

References

Asghar, A., & Libarkin, J. C. (2010). Gravity, magnetism and 'down': Non-physics college students' conceptions of gravity. *Science Education, 19*(1), 42-55.

Bani-Salameh, H. N. (2017a). Using the Method of Dominant Incorrect Answers with the FCI Test to Diagnose Misconceptions Held by First Year College Students. *Phys. Educ., 52*(2017a), 015006 (8pp).

Bani-Salameh, H. N. (2017b). How persistent are the misconceptions about force and motion held by College Students. *Phys. Educ., 52*(2017b), 014003 (7pp).

Bani-Salameh, H. N. (2017c). Teaching language effects on students' performance. *Health Professions Education, 4*(1), 27-30. https://doi.org/10.1016/j.hpe.2017.01

Bani-Salameh, H. N., Nuseirat, M., & Alkofahi, K. A. (2017a). Performance gap among male and female college students measured with the force concept inventory. *IOSR journal of applied physics, 9*(2), 11-13.

Bani-Salameh, H. N., Nuseirat, M., & Alkofahi, K. A. (2017b). Do first year college female and male students hold different misconceptions about force and motion? *IOSR journal of applied physics, 9*(2), 14-18.

Benson, J., & Clark, F. (1982). A guide for instrument development and validation. *American Journal Occupational Therapy, 36*(12), 789-800.

Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physics Review special topics physics Education Research, 5*, 020103-1-17.

Feeley, R. E. (2007). Identifying student concepts of gravity. *Masters thesis in Science and Teaching*, The University of Maine.

George, D., & Mallery, P. (2009). SPSS for Windows Step by Step: A Simple Guide and Reference. *Pearson Education,* Boston, MA.

Griethuijsen, R. A. L. F., Eijck, M. W., Haste, H., Brok, P. J., Skinner, N. C., & Mansour, N. et al., (2014). Global patterns in students' views of science and interest in science. *Research in Science Education, 45*(4), 581-603. https://doi.org/10.1007/s11165-014-9438-6

Gunstone, R. F., & White, R. T. (1981). Understanding of gravity. *Science Education, 65*, 291-299.

Haladyna, T. M., Downing S. M., & Rodriguez M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Appllied Measurement Education, 15*(3), 309-334.

Halloun, I., & Hestenes, D. (1985a). The Initial Knowledge State of College Physics Students. *American Journal of Physics, 53*(11), 1043.

Hambleton, R. K., & Jones, R. J. (1993). Comparison of classical test theory and item response theory and their application to test development. *Education Measurement: Issues Practice, 12*, 38-47.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *Physics Teacher, 30*, 141-158.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA.

Piburn, M. (1998). Misconceptions about gravity held by college students. Conference Proceedings, NARST Annual Meeting, Lake of the Ozarks, Missouri, April 10-13. *ERIC*

Document Reproduction Service, (292), 616, BBV1998.

Schlingman, W. M., Prather, E. E., Wallace, C. S., Rudolph, A. L., & Brissenden, G. (2012). A classical test theory analysis of the light and spectroscopy concept inventory national data set. *Astronomy Education Review, 11*, 010107.

Smith, C. L., & Treagust, D. F. (1988). Not understanding gravity limits students' comprehension of astronomy concepts. *The Australian Science Teachers' Journal, 33*, 21.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education, 48*, 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Wallace, C., & Bailey, J. M. (2010). Do concept inventories actually measure anything? *Astronomy Education Review, 9*, 010116.

Williamson, K. (2013). Development and calibration of a concept inventory to measure introductory college astronomy and physics students' understanding of Newtonian gravity. *Doctoral thesis,* Montana State University.

Williamson, K., & Willoughby, S. (2012). Student understanding of gravity in introductory college astronomy. *Astronomy Education Review, 11*, 010105.

Williamson, K., Prather, E., & Willoughby, S. (2016). Applicability of the Newtonian gravity concept inventory to introductory college physics classes". *American Journal of Physics, 84*(458). https://doi.org/10.1119/1.4945347

Williamson, K., Willoughby, S., & Prather, E. E. (2013). Development of the Newtonian gravity concept inventory. *Astronomy Education Review, 12*(1), 010107.

Copyright Disclaimer