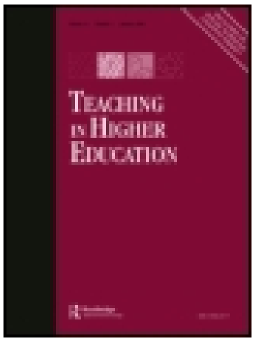

Teaching in Higher Education

Critical Perspectives

ISSN: 1356-2517 (Print) 1470-1294 (Online) Journal homepage: https://www.tandfonline.com/loi/cthe20

Analysing assessments in introductory physics using semantic gravity: refocussing on core concepts and context-dependence

Christine M. Steenkamp, Ilse Rootman-le Grange & Kristian K. Müller-Nedebock

TEACHING IN HIGHER EDUCATION
https://doi.org/10.1080/13562517.2019.1692335

Routledge
Taylor & Francis Group

Check for updates

Analysing assessments in introductory physics using semantic gravity: refocussing on core concepts and context-dependence

Christine M. Steenkamp [a], Ilse Rootman-le Grange [b] and Kristian K. Müller-Nedebock [a]

[a]Department of Physics, Stellenbosch University, Stellenbosch, South Africa; [b]Faculty of Science, Stellenbosch University, Stellenbosch, South Africa

ABSTRACT

The development of learning practices that enable students to transfer knowledge across contexts, is a dominant topic in Physics Education Research. Assessment is a key activity in the learning process. The purpose of this paper is to illustrate the value of analysing introductory physics assessments using the Semantics dimension of Legitimation Code Theory. We discuss the tools used to analyse the test and exam question papers of two consecutive calculus-based introductory physics modules. An analysis of past question papers over 5 years revealed various weaknesses. The outcomes of an intervention based on critical self-evaluation of question papers, using the same tools, are presented. The results indicate that the intervention increased focus on core concepts and context and supported learning that enables transfer. We argue that the use of semantic gravity to analyse assessments is a useful starting point for change in educational practices in order to support transfer.

ARTICLE HISTORY
Received 9 February 2019
Accepted 5 November 2019

KEYWORDS
Assessment; moderation; Legitimation Code Theory; Semantics dimension; introductory physics

Introduction

Physics is a discipline based on highly abstract core concepts. Each core concept is a generalised principle applicable to a class of phenomena that are found in a variety of real-world contexts. The core concepts are interrelated, building on one another to form a hierarchical knowledge structure (Bernstein and Solomon 1999). The core concepts are applied to explain and predict real-world phenomena and the ability to do so is critical to physicists on every level of education. It is expected from undergraduate physics students to be able to transfer their understanding of these core concepts across different real-world contexts, applying them in a range of commonplace to specialised situations. They must also transfer this knowledge over time to subsequent study years and related disciplines, where they are expected to cumulatively build new knowledge on these existing concepts.

This idea of transfer has been around for over 100 years, when psychologist Edward Thorndike was one of the first researchers to start investigating the processes involved

CONTACT Christine M. Steenkamp cmsteen@sun.ac.za Department of Physics, Stellenbosch University, Private Bag X1, Matieland 7602, Stellenbosch, South Africa

with transfer of knowledge and skills between different situations. (Nokes-Malach and Mestre 2013). Since then various definitions of transfer have been articulated and research on the subject has covered not only psychology, but has also moved to investigations in science disciplines. The classical perspective defines transfer as ‘the application of knowledge learned in one context to a new context’ (Lobato 2006). This is critically important in physics where it is required of students to be able to apply the core concepts across different parts of the curriculum and to real-life scenarios (Laverty et al. 2016). In a review by Docktor and Mestre (2014) exploring the major research areas in Physics Education Research (PER), research related to cognition is highlighted as one of the most prominent PER topics. Docktor and Mestre (2014) identify ‘learning and transfer’ as a subsection of cognition and specifically mentions aspects such as the impact that context has on knowledge transfer and different mechanisms of knowledge transfer. Examples of recent studies related to aspects of knowledge transfer in PER include that of Podolefsky and Finkelstein (2006, 2007a, 2007b) that specifically look at the role that analogies play in supporting knowledge transfer, as well as Brookes, Ross, and Mestre (2011) who investigate the impact of the specificity effect on transfer.

Yet, despite research on this topic stretching over roughly 100 years, Barnett and Ceci reported in 2002 that researchers still hold contradicting views (Barnett and Ceci 2002, 612). They are supported by calls from Lobato (2006) and Schoenfeld (1999) for the development of new theories of transfer to support research in this area. Therefore, in this study, we explore Legitimation Code Theory (LCT) (Maton 2014b) as an alternative theoretical framework for analysing the classical transfer approach in a physics education context. Transfer is a key concept in LCT. It is incorporated into the term ‘cumulative learning’ (e.g. Maton 2014b) that is defined as a learning process that results in students being ‘able to transfer knowledge across contexts and through time’ (Maton 2009). The term ‘cumulative learning’ highlights that transfer is essential when students must build new knowledge on previously acquired knowledge, as required for the highly hierarchical knowledge structure of Physics in particular (Maton 2009). In this study, we will focus on one aspect of LCT, namely the concept of semantic gravity, which explores the impact that context has on knowledge transfer, resonating with the aspects that are highlighted by the review of Docktor and Mestre (2014).

Problem statement and aims

In first-year physics cumulative learning means that students know (have understood and memorised) the core concepts in each section of the curriculum and are able to transfer this knowledge to analyse phenomena and solve problems in different real-world contexts by identifying the relevant core concepts and applying these (either conceptually or by calculation). Our observation is that many first-year students are unfamiliar with this approach to learning and struggle to acquire it. For example, in written assessments, it is often observed that students cannot correctly formulate core concepts. Furthermore, they struggle to identify the relevant core concepts when challenged with solving unfamiliar problems and instead try to re-apply the patterns of previously worked-out, similar examples. Reasons for this challenge can be sought in the school curriculum or in the pressure on schools to train learners for optimum performance in the National Senior Certificate exams. However, we suggest that a more constructive approach is to ask to what extent our educational practices in undergraduate, and specifically first-year university physics, cultivates cumulative learning – defined as a learning process that facilitates transfer.

This study focuses on analysing summative assessment questions in two consecutive introductory undergraduate physics modules to investigate the effect on transfer. The first reason for the study is that assessments concretely communicate to students what is expected of them and always functions as an intervention into student learning (Bearman et al. 2016; Brown and Knight 1994). The second reason is pragmatic. The challenge is to develop cumulative learning in a real academic environment, where the lecturers and teaching styles may change from one year to the next. It is, however, possible to make changes in the process of formal assessment (tests and exams) since this is a regulated process involving all the lecturers and an internal moderator. The expectation was that sensitising lecturers by means of the analysis of assessments may also influence their approach to teaching.

An internal moderation process is implemented at our institution. In the case of first-year physics modules each test or exam question paper, together with the model answers, is reviewed by an internal moderator who has relevant experience but is not involved in teaching the module. In the past, however, it was typical that the questions in a question paper did not change much during this moderation process. We did not have a framework for discussing the difficulty of questions in order to develop a shared understanding of standards (Beutel, Adie, and Lloyd 2017). This resulted in varying judgements, as shown before (Fakcharoenphol, Morphew, and Mestre 2015). Arguments on the difficulty of questions were often based on whether that particular type of problem has been discussed in class, thus ignoring that we do not want students to memorise worked-out examples. The result in the past was inconsistency in the difficulty of assessments, which resulted in widely varying average class marks.

The first aim of this study was to determine to which extent the Semantics dimension of LCT is a useful tool to analyse assessment questions in first-year physics modules and whether a correlation between this analysis and student marks exist. The second aim was to use semantic gravity to analyse and self-critique new question papers during the moderation process and evaluate the impact of this intervention on the learning process and students' ability to transfer knowledge.

The Semantics dimension of Legitimation Code Theory

LCT provides a theoretical framework for analysis of educational practices, as developed by Maton and co-workers. It forms part of the philosophy of social realism, where knowledge is considered to be produced socially and have real effects (Georgiou, Maton, and Sharma 2014). It offers a suite of tools to analyse educational practices (Maton 2014b, 17–19; Maton and Moore 2010) and is used increasingly as framework for empirical research into educational practices. LCT builds on the ideas of Bernstein (Bernstein and Solomon 1999) of the distinction between horizontal and hierarchical knowledge structures. In physics the hierarchical knowledge structure, where new knowledge builds on existing knowledge, is applicable.

The Semantics dimension is one of the tools of LCT. This dimension offers a distinction between context-dependence and complexity that has proved useful in analysis of educational practices (Maton 2014a). Examples of its application in the South-African tertiary education system have been reported in the broad fields of education (Shay 2008), physics (Conana, Marshall, and Case 2016), chemistry (Blackie 2014; Rootman-le Grange and Blackie 2018), biology (Kelly-Laubser and Luckett 2016; Mouton and Archer 2018), journalism (Kilpert and Shay 2013) and political science (Clarence 2016). Context-dependence refers to the degree to which the meaning of a term or concept is related to a specific context, and in LCT this is termed semantic gravity (SG). Complexity refers to the degree to which meaning is condensed into specialised terms and symbols and the term sematic density (SD) is used to describe it.

In LCT, it is convention to construct a Semantics plane with semantic density and semantic gravity on the x and y axes of the Cartesian plane, respectively, and to represent categories of knowledge as points on the plane. To analyse academic content using this representation it is important to first clearly define the context to which the semantic gravity relates. Secondly, a translation device must be developed for each of the parameters, to define the scale along the individual axes. It has to be stated explicitly that neither semantic gravity nor semantic density can be categorised into absolute strengths, rather it is relative strengths on a continuous scale that offer unlimited levels of differentiation, which is determined by the problem being investigated (Georgiou, Maton, and Sharma 2014). In some instances, it is useful to describe practices by semantic gravity ranges (the difference between the strongest and weakest semantic gravity present) rather than semantic gravity strengths (Maton 2014b).

Shay (2008) has prepared the theoretical ground for application of LCT to assessment, concluding that LCT may be a useful theoretical framework for conceptualising the relation between knowledge and evaluative criteria used in assessments. In PER semantic gravity has been used to analyse students' responses to a thermal physics question (Georgiou, Maton, and Sharma 2014) and to profile students' methods of problem-solving (Conana, Marshall, and Case 2016). In both these cases, the students' answers were analysed. In this work, we apply LCT to analysis of assessment questions in physics (not the student answers), as has been done in chemistry (Blackie 2014; Rootman-le Grange and Blackie 2018), but not yet to our knowledge in physics. Laverty and co-workers recently developed a protocol for analysis of assessment questions (Laverty et al. 2016) and applied it to the standardised conceptual assessments (concept inventories) that are available for Physics (Laverty and Caballero 2018). This analysis distinguishes between tasks assessing 'core concepts', 'scientific practice' (constructing models and explanations) and 'crosscutting concepts' that are described as 'ways of thinking' common to all scientific disciplines. Although this analysis protocol shows similarity to semantic gravity analysis and has the same aim – to enable students to apply their knowledge across different contexts – it is not linked to LCT theoretically.

Our study was motivated by conclusions of previous research which argued that cumulative learning can only take place when students are systematically exposed to context-independent as well as context-dependent knowledge, i.e. to a range of semantic gravity levels (Maton 2013; Georgiou, Maton, and Sharma 2014) and that the semantic gravity range represented in assessments is an important factor in facilitating cumulative learning (Kilpert and Shay 2013). Rootman-le Grange and Blackie (2018) have shown that the limited variation in context-dependence (semantic gravity) in chemistry assessments could be the reason that students struggle to transfer their knowledge between different contexts (Rootman-le Grange and Blackie 2018). We, therefore, expected that mapping of the semantic gravity levels and ranges present in our assessments would likewise shed light on our challenges in first-year physics.

Methodology

We applied semantic gravity to analyse assessment questions of two calculus-based introductory physic modules. The modules are the first and second-semester physics modules (referred to for simplicity as Ph101 and Ph102 in this paper) in the first-year of a three-year Bachelor of Sciences degree programme. Most students enter the first-semester module directly after graduating from high school. Ph101 and Ph102 are compulsory for students enrolled for programmes in experimental and theoretical physics (approximately 12% of class in the past 2 years), mathematics, applied mathematics, computer science, earth sciences, geoinformatics, chemistry and polymer science and chemical biology. It is an elective module for the programmes in molecular biology and biotechnology, as well as human life sciences. In Ph101 and Ph102 students are expected to build a knowledge base of core physics concepts and develop problem-solving skills that will facilitate their success in their respective programmes.

The two modules constitute a traditional calculus-based introduction to physics, consisting of Newtonian mechanics, gravity, fluid mechanics and thermodynamics in the first semester, followed by electricity, magnetism and special relativity as main topics in the second semester. The textbook that is used is 'Sears and Zemansky's University Physics' authored by Young and Freedman (Young and Freedman 2016).

In the development of the translation device for this study, the following requirements were important. Firstly, the translation device must be useful for all types of questions in our assessments. Secondly, the semantic gravity scale must reflect the degree to which the students are required to transfer knowledge from the curriculum to curriculum-independent scenarios, as this is an important quantity to measure. Thirdly, the semantic gravity scale should accommodate the tension between the abstract context of the curriculum and the real-world context, which is usually interpreted as stronger semantic gravity. As result core concepts that are strongly curriculum dependent are coded as weaker semantic gravity, whereas real-life applications of such concepts are coded as stronger semantic gravity. Finally, the device must be clear and useful to physics colleagues who are not LCT experts.

The translation device that we used to characterise relevant semantic gravity strengths is presented in Table 1. It defines four strengths, labelled SG−−, SG−, SG+ and SG++ in order from weaker to stronger semantic gravity. The weaker SG−− level represents the formulation of a core concept without reference to a concrete example or application. The SG− level represents the application of a core concept to a simplified scenario that is associated, by design, to the section of the curriculum where this concept was discussed. The SG+ level represents more complex empirical scenarios, closer to real-world problems, but still well-defined and associated with at least one of the contexts discussed in the curriculum. The SG++ level represents real-world grounded scenarios that require everyday experience and logical reasoning in addition to curriculum content to solve the problem or explain the phenomenon.

Table 1. Translation device used for semantic gravity levels.

Level	Criteria
SG++	Application of a core concept to a 'real world' problem that can only be solved by additionally applying everyday knowledge and self-defined assumptions.
SG+	Application of a core concept to a well-defined empirical scenario, where the association of the core concept with the scenario has not been discussed in the curriculum, although the scenario lies within the scope of the curriculum.
SG−	Application of a core concept to a simplified empirical scenario that is associated with this specific core concept in the curriculum.
SG−−	Formulation of a core concept (a general principle, concept, definition or law) that is found in one clearly defined section of the curriculum, without reference to an example from the empirical domain.

As in the pioneering study of Georgiou, Maton, and Sharma (2014), we did not use semantic gravity strengths, but rather semantic gravity ranges to categorise the assessment questions, asking what semantic gravity range is required to answer each question successfully. The semantic gravity ranges are presented in Table 2. Since all questions require the identification and formulation of core concepts, the weakest semantic gravity of every range is SG−−. The semantic gravity ranges do however extend to different semantic gravity levels as defined in the translation device (Table 1): SGR1 means that only SG−− is included, whereas SGR2 ranges from SG−− to SG−, SGR3 from SG−− to SG+ and SRG4 from SG−− to SG++. It should be noted that the number used to label each of the semantic gravity ranges indicates the reach of the range over levels of relative strengths and should not be associated with an absolute value. This semantic gravity range device indicates the degree to which transfer of knowledge is required, from little transfer in SGR1 questions to a high degree of transfer to a scenario outside the curriculum in SGR4 questions.

The test and exam question papers of Ph101 and Ph102 seldom contain multiple-choice questions, and if so, such questions only make up a small fraction of the marks. Most of the questions require an answer in prose, or a calculation. A question can be a single statement, can consist of a grouping of a few sub-questions on similar topics or sub-questions leading through successive steps. Examples of questions from the different semantic gravity ranges are presented in Table 3. The semantic gravity range of each numbered sub-question was classified, as the first step. A representation of each assessment, showing the total number of marks allocated to each semantic gravity range was produced. This representation is termed the semantic gravity range map of the assessment. The average sematic gravity range of a question (that consists of sub-

Table 2. Definition of semantic gravity ranges used for in the analysis of assessment questions.

Range	Strengths included	Criteria
SGR4	SG−− to SG++	The question requires the integration of core concepts (SG−−), possibly from different sections of the curriculum (SG+), with everyday knowledge and self-defined assumptions to solve a 'real world' problem outside the context of the curriculum (SG++).
SGR3	SG−− to SG+	The question requires the integration of core concepts (SG−−) from different sections of the curriculum in order to solve a problem with a wider context (SG+).
SGR2	SG−− to SG−	The question requires the application of one concept (or a few closely related concepts, SG−−) from one clearly defined section of the curriculum to a specific empirical example within the context of that section of the curriculum (SG−).
SGR1	SG−−	The question requires the identification and formulation of a concept (a concept, definition or rule) that is found in one clearly defined section of the curriculum.

Table 3. Examples of questions from different semantic gravity ranges.

Range	Example question	Explanatory notes
SGR4	Water has the important property that it has the highest density at 4 degrees Celsius. Explain why this property is important to prevent large water bodies (like lakes) from freezing solid down to the bottom in cold places where the air temperature is below freezing temperature in winter.	The concepts of density, heat transfer and phase change must be used, and combined with general knowledge to argue: If air is cold it means the lake loses heat from its surface – what happens then?
SGR3	A hot air balloon floats stationary in the air. The mass of the balloon is 200 kg and its volume is 1000 m^3. The temperature of the surrounding cold air is 10° Celsius and the atmospheric pressure is 101,300 Pa. Treat air as an ideal gas. Calculate the density of the hot air in the balloon.	The concepts of forces in equilibrium (from mechanics), the properties of an ideal gas (from the thermodynamics section) and the definition of density must be integrated.
SGR2	The absolute pressure experienced at the bottom of a freshwater lake is 248,300 Pa. Calculate the depth of the lake. The atmospheric pressure is 101,300 Pa.	The concept of pressure in a fluid is needed to solve the problem.
SGR1	(i) Explain what is meant by the triple point of water. (ii) Define what is meant by a conservative force.	In each case one core concept must be recalled and defined or explained.

questions) was calculated by a weighted combination of the sub-questions, where the contribution of each sub-question to the average semantic gravity range is weighted by the number of marks the sub-question counts towards the question total. In a similar way the average semantic gravity range of the complete assessment is calculated by combining the different questions.

Our investigation started with a historic analysis of the test and exam question papers of P101 and Ph102 over a period of 5 years (2012–2016). We analysed the question papers of the major test that is scheduled about two thirds into each semester and the first exam of every semester module, as this exam is written by most of the students. This analysis was done by one researcher (CMS). The average marks shown for the historic study represent the official statistics, including students who were enrolled for the first time and students who repeated the module.

The second part of the study consisted of an intervention during the 2017 academic year. Analysis of every new question paper was done during interactive internal moderation sessions. The lecturers of the module and the internal moderator were present. The semantic gravity ranges were determined by consensus. A goal was set for the semantic gravity range map of the paper, for example in Ph101 the goal was to assign approximately 50% of the marks to SGR1 and SGR2 questions, between 10% and 20% of marks to SGR4 questions, and the remaining marks to SGR3 questions. Changes were made to questions in order to adjust the semantic gravity range map of the question paper to reach the goal. We communicated early and clearly to the students that assessments will contain 50% of questions on core concepts and their direct applications (SGR1 and SGR2).

The only other change that was made during 2017 in comparison to previous years was that mathematical methods were taught when needed instead of giving students an introduction to all the required mathematics at the beginning of the semester. The format of our lectures and tutorial sessions did not change significantly and the range of problems given in the tutorials were as usual of the SGR2, SGR3 and SGR4 ranges in order to develop the students' problem-solving skills.

The students' performance before (2016) and after the intervention (2017) were compared for the group of students who registered for undergraduate studies for the first time during each year. This selection was made to avoid the hysteresis effect of previous years on the results when including students who are repeating the module. The small number of students who deregistered for the modules before the first test, or who did not register for Ph102 despite having passed Ph101, were not included in the statistics. In addition, a detailed analysis was done of the students' answers to the first exam of Ph101 in 2016 and 2017, respectively, to determine the average mark of the cohort in questions of a particular semantic gravity range. This was done for a sample of 50% of each cohort, selecting every second name on the alphabetical list, due to the size of the cohorts and the assumption that 50% is a representative sample.

Results and discussion

Study of historic papers

The results on question papers for the period 2012–2016 are presented in Figure 1. Figure 1(a,b) shows the semantic gravity range maps of question papers for Ph101 and Ph102, respectively. The percentage of the marks in the question paper that was allocated to questions of different semantic gravity ranges are presented as a bar graph. In Figure 1(c,d), the cohort's average mark (in percentage) for the corresponding question papers for Ph101 and Ph102, respectively are presented.

Figure 1 reveals significant variation in both the semantic gravity range maps of the assessments and the average marks. The average marks for exams vary in the range 33%–78% in Ph101 and 28%–59% in Ph102. Although these are not the only marks that are used to calculate students' final marks there was reason for concern. In the semantic gravity range maps the percentages of the questions that are in the SGR1 and SGR2 categories of weaker semantic gravity curriculum-bound questions vary in the range of 0%–68% for Ph101 and 21%–57% for Ph102. In five question papers, the strongest semantic gravity range category SGR4 is dominant (> 40%). In five question papers no questions of the SGR1 category, which directly assess the core concepts, have been asked. It must be concluded that although we consider the core concepts very important, direct assessment of these have a relatively low mark weight in our assessments. This may lead students to underestimate the importance of knowing the basics.

It can be seen from Figure 1 already that there is a degree of correlation between the semantic gravity range maps of the assessments and the average marks. A curve was drawn in Figure 1(a,b) to indicate the percentage of questions in the SGR1 and SGR2 categories. In three cases where the papers contained no significant SGR4 questions the curve was adjusted slightly upwards to compensate. The resulting curve shows similar trends to a curve drawn through the average marks in Figure 1(c,d).

A quantitative correlation between marks and the semantic gravity range maps was done by plotting the average marks obtained in the tests and first exams of 2012–2016 against the average semantic gravity ranges of the question papers. The average semantic gravity was quantified by assigning values 1, 2, 3, 4 to semantic gravity range categories SGR1, SGR2, SGR3 and SGR4 and calculating a weighted average. For example, a paper with no SGR4 questions and an equal representation of the other semantic gravity ranges will have an

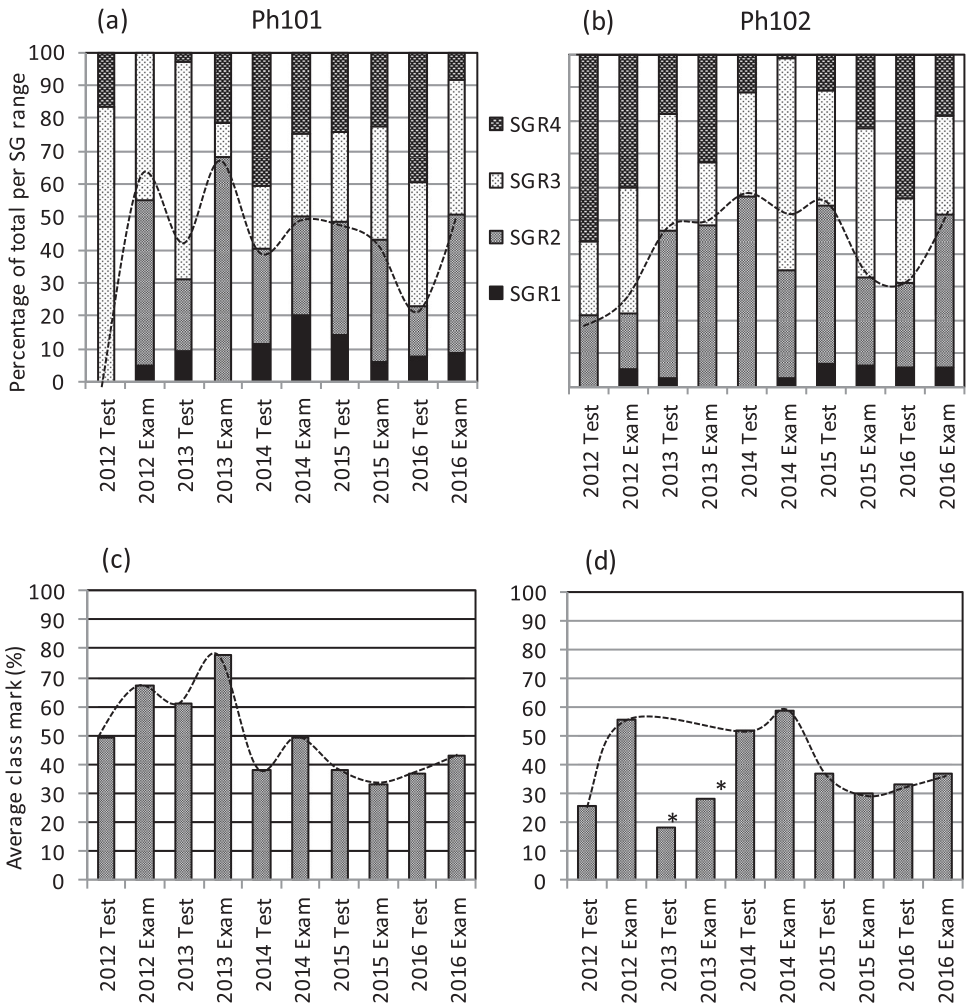

Figure 1. Graphs (a) and (b) represent the semantic gravity range maps of test (T) and exam (E) papers in Ph101 and Ph102, respectively. In graphs (c) and (d) the average class marks for the corresponding assessments are plotted. Curves are drawn to highlight trends. Note: During 2013 problems were experienced in Ph102 due to special circumstances and the 2013 average marks indicated by * should not be considered part of the general trend.

average of 2, a paper with equal representation of all the semantic gravity ranges would have an average of 2.5, and a paper with no SGR1 questions but equal representation of the other semantic gravity ranges will have average of 3. The results are plotted in Figure 2(a,b) for Ph101 and Ph102, respectively. The data shows that most of our 2012–2016 assessments have average semantic gravity above 2.5. The plots in Figure 2 show downward slopes meaning that the average mark decreases as the average semantic gravity of the question paper is strengthened. The only exception is the final exam at the end of the year in Ph102 where the data are scattered, and the trend is unclear. The decrease of the average mark with strengthening of semantic gravity is more pronounced during the first semester (Ph101) than during the second semester (Ph102).

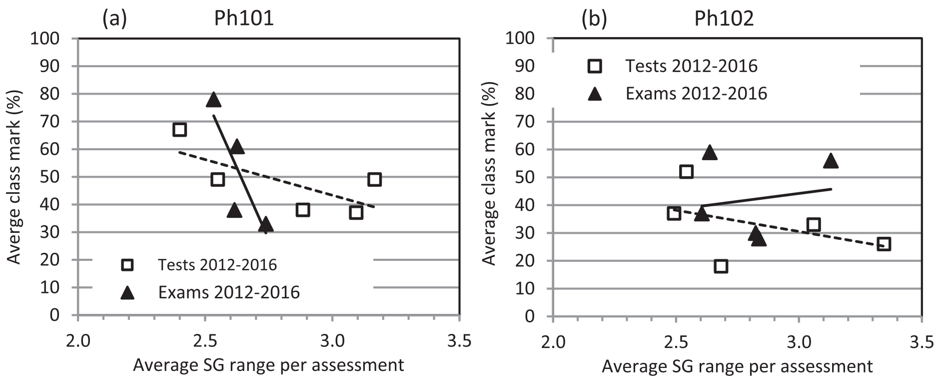

Figure 2. The average class marks for the main test and first exam over 5 years (2012–2016) are graphed against the average semantic gravity range of the question paper. The average semantic gravity range was calculated by assigning values 1, 2, 3, 4 to semantic gravity ranges SGR1, SGR2, SGR3 and SGR4, respectively and calculating a weighted average.

These results demonstrate firstly that students struggle to transfer their knowledge to different contexts within and outside the curriculum. Secondly, the results show that the semantic gravity range maps of assessments do have an influence on the marks and should, therefore, be managed in the interest of consistent and fair assessments. This is particularly critical in the first semester.

From 2015 onwards the Ph101 marks for individual questions were archived, and assessments were analysed on the level of individual questions. In Figure 3 the average mark of the cohort for every question is plotted against the average semantic gravity range of the question. In case of a compound question the average semantic gravity range of the question was calculated. The average mark does not show a clear decrease with strengthening of the semantic gravity range of the question, but a maximum with a decreasing trend towards both weaker and stronger semantic gravity. This shows that the students are best prepared for questions of intermediate semantic gravity – typically problems involving core concepts from one or more sections of the textbook in a calculation. It confirms the observation by the lecturers that students do not answer direct

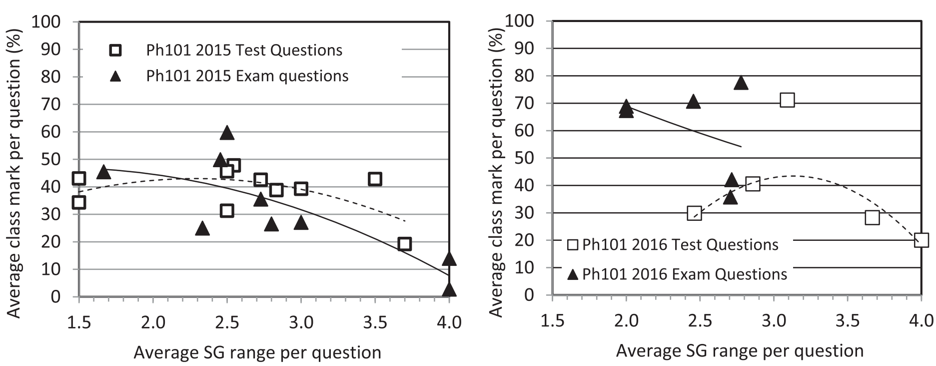

Figure 3. The average mark of the class per question is graphed against the average semantic gravity range of the individual questions for the class tests and exams of Ph101 in 2015 (a) and 2016 (b).

questions on core concepts (SGR1) as well as expected. Students also struggle with questions of stronger semantic gravity (SGR4) showing that their ability to transfer knowledge to everyday situations is limited. This shows that in their preparation for assessments students are most probably concentrating on the type of problem-solving demonstrated during lectures and in the end-of-chapter textbook problems and that we fail to transfer the message that knowing and understanding the core concepts is most important.

We concluded from these results that the semantic gravity range maps of the question papers in Ph101 and Ph102 varied significantly in the historic papers. This variation has an influence on student success, especially in the first semester, and large variation hinders consistent and fair assessment practice.

The weakness in our assessments differs from that identified in the recent analysis of chemistry assessments in our institution (Rootman-le Grange and Blackie 2018). The chemistry assessments were found not to test a sufficiently large range of semantic gravity, in particular lacking real-world problems, therefore not encouraging students to apply their knowledge in different contexts. Our past assessments tested a large range of context-dependence, but with underrepresentation of questions about the core concepts in the SGR1 category. We consider knowledge of the core concepts crucial to enable students to transfer their knowledge to new contexts, but our assignments did not encourage students to study the core concepts. The question-level analysis shows that this results in students struggling both at the weaker and at the stronger semantic gravity ends of the range. An intervention is needed to encourage learning that supports knowledge transfer by assessing more consistently over the whole semantic gravity range that we want the students to master, including direct assessment of core concepts in order to emphasise their importance.

Intervention during 2017

Analysis and control of the semantic gravity range maps of question papers were performed during the internal moderation process. This intervention changed the dynamics of the internal moderation process significantly. The use of semantic gravity to analyse the assessments was empowering to the lecturers and internal moderators. For the first time, we had a clear and practical way to categorise the questions in a paper. The result was that the internal moderation lead to much more significant changes to the question papers than had been typical in the past. Furthermore, the discussions forced the lecturers to clarify to themselves what the core concepts are. The in-depth discussion of question papers also resulted in a heightened awareness, amongst lecturers, of the diverse real-life experiences ('contexts') of our students and the possible misunderstanding of questions that this may cause, especially in the stronger semantic gravity ranges.

The focus on core concepts, which was instrumental in analysing the semantic gravity of the assessment questions, had a knock-on effect in teaching. Lecturers were more aware of what aspects of the curriculum were core concepts and placed more emphasis on these and on identifying which concepts are used in applications.

The most significant qualitative observation concerning student behaviour is that the pre-exam 'panic' among the students, that has occurred in Ph101 in most previous years, was absent during 2017. This is in reference to the typical increase in requests for consultations with lecturers shortly before the exam, questions about how to prepare and requests for copies of previous exam question papers (which are never provided).

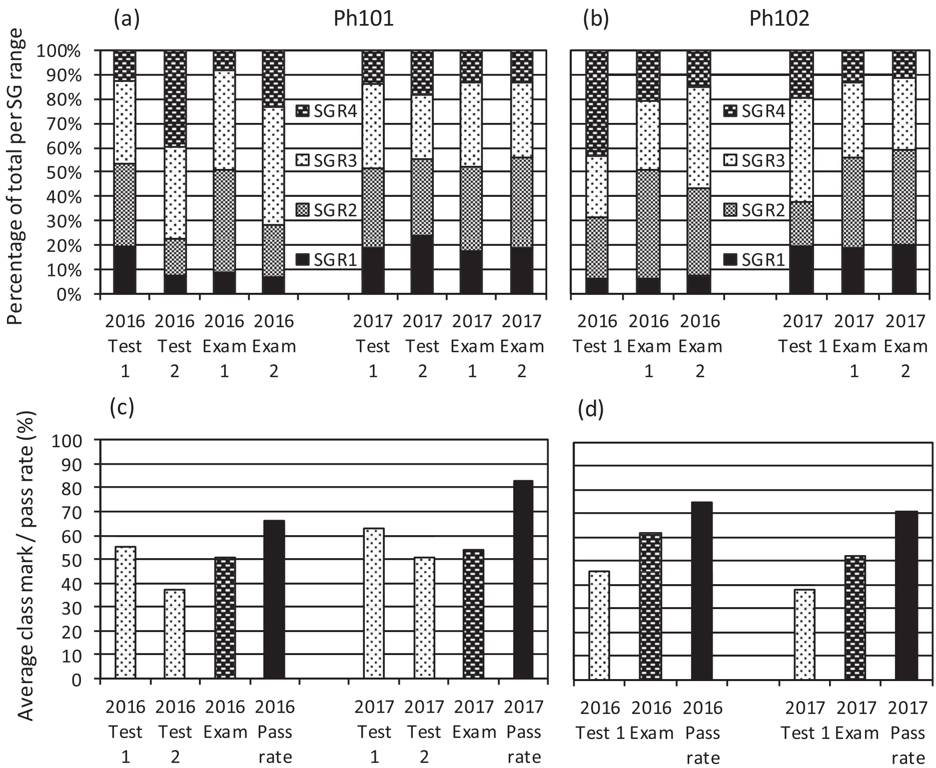

Figure 4. The semantic gravity range maps of assessments before (2016) and during the intervention (2017) are shown for Ph101 (a) and Ph102 (b) The average marks for first time registered students who enrolled for both modules are shown in (c) and (d), respectively.

In Figure 4 the assessments of Ph101 and Ph102 during 2016 (before the intervention) and 2017 (during the intervention) are compared. Only the students who had registered for the first time at the university that year (approximately 67% of the class) were included in the statistics.

In Ph101, the semantic gravity range maps of the assessments during 2017 were (by design) more consistent than in 2016, correlating with less variation in the average marks and a higher pass rate (refer to Figure 4(a,c)). The variation in marks during 2016 can be rationalised in terms of the semantic gravity range maps as the second test had a much stronger average semantic gravity and therefore lower average marks. It can also be seen that in 2016 the semantic gravity of the second exam was significantly stronger than that of the first exam, to the disadvantage of students who chose to write the second exam. The pass rate of the first-time registered students for Ph101 in 2017 was 84% – a significant improvement compared to 66% in 2016. The decision to strictly control the semantic gravity range maps of assessments in Ph101 (during the first semester) was informed from the historic study that showed that the semantic gravity range map has a more pronounced effect on marks in the first semester than in the second. This intervention in Ph101 had impacted positively on the focus of lecturers and students on core concepts. It also positively affected the average marks and the pass rate of Ph101.

In Ph102, the same semantic gravity analysis was done, but the semantic gravity range map was not controlled as strictly (Figure 4(b)). A reduced influence of the intervention on the marks and pass rate is therefore expected (Figure 4 (d)). In both 2016 and 2017, the test had stronger semantic gravity than the exams, correlating with a lower average mark in the test. The pass rate of the first-time registered students in Ph102 in 2017 was slightly (4%) higher than in 2016. The important message in the result is that the students who progressed from Ph101 to Ph102 during the intervention in 2017 were at least as well prepared for the second semester as in previous years.

A detailed analysis of the students' answers to the first exams of Ph101 in 2016 and 2017 was done to determine how the students performed in questions of the respective semantic gravity range categories. The results are shown in Figure 5.

In the 2016 results, before the intervention, Figure 5 displays the typical trend that students do not perform best in SGR1 questions. That means they either could not recall the relevant core concepts at all, or could not identify the correct core concept, or could not formulate or explain it correctly. During the intervention of 2017, this changed and the SGR1 category showed the highest average, indicating that our attempt to encourage students to know the core concepts well has been successful. In the SGR4 category the marks also improved significantly. This may be evidence that being more aware of what the core concepts are and being able to formulate them correctly facilitates the ability of the students to apply these concepts in 'real world' scenarios that are not closely related to the curriculum context. However, in the SGR3 category there is a slight reduction in marks, possibly because students did not focus so much on practicing these as in the past. There was no significant change in the SGR2 category.

The analysis of average marks per semantic gravity range confirms that the intervention improved students' focus on and therefore knowledge of the core concepts. This agrees with anecdotal evidence from lecturers about an increased focus on core concepts, as a result of the intervention. To answer the question whether the intervention encouraged cumulative learning different factors should be considered. The increase in average marks in the SGR4 category suggest an improvement in the ability to transfer knowledge

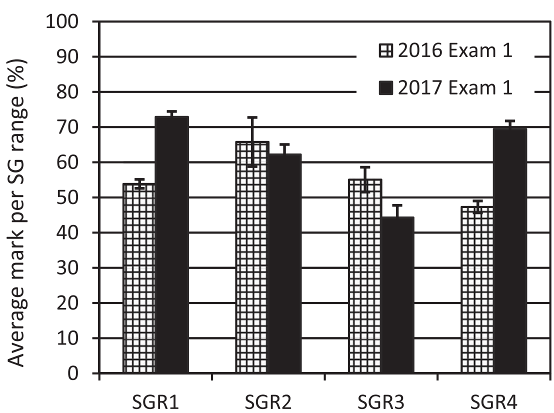

Figure 5. The students' answers to the first exam of Ph101 before (2016) and during the intervention (2017) were analysed on question-subsection level to determine the average mark in all questions of a particular semantic gravity range category.

across contexts, but the decrease in marks in the SGR3 category and the marginal improvement in the marks in the second semester serve as reminder that the development of cumulative learning is not achieved in one semester, but that a continued effort is needed.

Conclusions towards future work

We conclude that we were able to identify and correct a weakness in introductory physics assessments using semantic gravity analysis. Underrepresentation of questions about the core concepts in our past assessments has served as miscommunication about the importance of the core concepts to students. It may have encouraged the students to focus on examples of applications in their preparation for assessments rather than on mastering the core concepts, leading to an approach of trying to adapt previous worked-out examples to solve unfamiliar problems, rather than identifying the core concepts which the problems have in common (Docktor and Mestre 2014).

We argue that the analysis of assessments using semantic gravity is a useful starting point for changing educational practice. Assessment is a high stakes activity for both lecturers and students and therefore a change in assessment practice has a direct effect on the learning-focus, as demonstrated in this study. It is a convenient starting point, as the setting and moderation of test and exam question papers is generally the most regulated part of the educational practice. In our experience, the relatively small effort of analysing assessments did not only lead to more effective moderation of assessments, but also to an increased awareness of what the core concepts are among both lecturers and students. It also yielded an increased awareness among the lecturers of the effect that different real-life experiences ('contexts') of students may have on their learning. We have continued to analyse our assessment questions to date with consistent results.

The Semantics dimension of LCT is highly useful to disciplinary experts (physicists in this case) as it puts the disciplinary knowledge in the centre (Maton 2013; Clarence 2016). Discussing semantic gravity categories immediately involve disciplinary knowledge and challenge the lecturers to rethink the dependence of meaning on context. Our experience of semantic gravity as being empowering agrees with the formulation by Clarence (2016, 135).

> Most exciting for the educators I have worked with thus far is that these LCT tools offer them a productive language with which to 'speak' their disciplines, as they also speak about recontextualisation of knowledge into curriculum, or the enactment of building that knowledge with their students into relevant systems of meaning through pedagogy and assessment.

Future work may include an effort to establish a useful translation device for semantic density in introductory physics assessment questions. Whereas semantic gravity on its own has proven to be a useful and empowering tool, in terms of rethinking the focus and fairness of our assessments, semantic density might enable us to shed more light on how students' different real-life experiences may influence their learning experiences.

The use of semantic gravity for assessment analysis has not (yet) been applied to analyse the structure of our lectures or tutorials, but an awareness has been established among lecturers. The results of this study on assessment can be used to guide changes to instruction and curriculum design (Maton 2013; Blackie 2014; Conana, Marshall, and Case 2016).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research has been funded as part of a grant from the Fund for Innovation and Research into Learning and Teaching (FIRLT) of Stellenbosch University.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author subject to institutional permission. The data are not publicly available due to restrictions of institutional rules and privacy protection principles.

ORCID

Christine M. Steenkamp http://orcid.org/0000-0003-0323-3793
Ilse Rootman-le Grange http://orcid.org/0000-0001-9799-7553
Kristian K. Müller-Nedebock http://orcid.org/0000-0002-1772-1504

References

Barnett, S., and S. J. Ceci. 2002. "When and Where do We Apply What We Learn? A Taxonomy for Far Transfer." *Psychological Bulletin* 128 (4): 612–637. doi:10.1037//0033-2909.128.4.612.

Bearman, M., P. Dawson, D. Boud, S. Bennett, M. Hall, and E. Molloy. 2016. "Support for Assessment Practice: Developing the Assessment Design Decisions Framework." *Teaching in Higher Education* 21 (5): 545–556. doi:10.1080/13562517.2016.1160217.

Bernstein, B., and J. Solomon. 1999. " 'Pedagogy, Identity and the Construction of a Theory of Symbolic Control': Basil Bernstein Questioned by Joseph Solomon." *British Journal of Sociology of Education* 20 (2): 265–279. doi:10.1080/01425699995443.

Beutel, D., L. Adie, and M. Lloyd. 2017. "Assessment Moderation in an Australian Context: Processes, Practices, and Challenges." *Teaching in Higher Education* 22 (1): 1–14. doi:10.1080/13562517.2016.1213232.

Blackie, M. A. L. 2014. "Creating Semantic Waves: Using Legitimation Code Theory as a Tool to Aid the Teaching of Chemistry." *Chemistry Education Research and Practice* 15 (4): 462–469. doi:10.1039/C4RP00147H.

Brookes, D. T., B. H. Ross, and J. P. Mestre. 2011. "Specificity, Transfer, and the Development of Expertise." *Physical Review Special Topics – Physics Education Research* 7 (1): 010105. doi:10.1103/PhysRevSTPER.7.010105.

Brown, S., and P. Knight. 1994. *Assessing Learners in Higher Education*. London: Kogan Page.

Clarence, S. 2016. "Exploring the Nature of Disciplinary Teaching and Learning Using Legitimation Code Theory Semantics." *Teaching in Higher Education* 21 (2): 123–137. doi:10.1080/13562517.2015.1115972.

Conana, H., D. Marshall, and J. Case. 2016. "Exploring Pedagogical Possibilities for Transformative Approaches to Academic Literacies in Undergraduate Physics." *Critical Studies in Teaching and Learning* 4 (2): 28–44. doi:10.14426/cristal.v4i2.73.

Docktor, J. L., and J. P. Mestre. 2014. "Synthesis of Discipline-Based Education Research in Physics." *Physical Review Special Topics – Physics Education Research* 10 (2): 020119. doi:10.1103/PhysRevSTPER.10.020119.

Fakcharoenphol, W., J. W. Morphew, and J. P. Mestre. 2015. "Judgements of Physics Problem Difficulty among Experts and Novices." *Physical Review Special Topics – Physics Education Research* 11 (2): 020128. doi:10.1103/PhysRevSTPER.11.020128.

Georgiou, H., K. Maton, and M. Sharma. 2014. “Recovering Knowledge for Science Education Research: Exploring the “Icarus Effect” in Student Work.” *Canadian Journal of Science, Mathematics and Technology Education* 14 (3): 252–268. doi:10.1080/14926156.2014.935526.

Kelly-Laubser, R. F., and K. Luckett. 2016. “Differences in Curriculum Structure Between High School and University Biology: The Implications for Epistemological Access.” *Journal of Biological Education* 50 (4): 425–441. doi:10.1080/00219266.2016.1138991.

Kilpert, L., and S. Shay. 2013. “Kindling Fires: Examining the Potential for Cumulative Learning in a Journalism Curriculum.” *Teaching in Higher Education* 18 (1): 40–52. doi:10.1080/13562517.2012.678326.

Laverty, J. T., and M. D. Caballero. 2018. “Analysis of the Most Common Concept Inventories in Physics: What Are We Assessing?” *Physical Review Physics Education Research* 14: 010123. doi:10.1103/PhysRevPhysEducRes.14.010123.

Laverty, J. T., S. M. Underwood, R. L. Matz, L. A. Posey, J. H. Carmel, M. D. Caballero, C. L. Fata-Hartley, et al. 2016. “Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol.” *PLOS ONE* 11 (9): e0162333. doi:10.1371/journal.pone.0162333.

Lobato, J. 2006. “Alternative Perspectives on the Transfer of Learning: History, Issues, and Challenges for Future Research.” *The Journal of the Learning Sciences* 15 (4): 431–449. doi:10.1207/s15327809jls1504_1.

Maton, K. 2009. “Cumulative and Segmented Learning: Exploring the Role of Curriculum Structures in Knowledge-Building.” *British Journal of Sociology of Education* 30 (1): 43–57. doi:10.1080/01425690802514342.

Maton, K. 2013. “Making Semantic Waves: A Key to Cumulative Knowledge Building.” *Linguistics and Education* 24 (1): 8–22. doi:10.1016/j.linged.2012.11.005.

Maton, K. 2014a. “A TALL Order? Legitimation Code Theory for Academic Language and Learning.” *Journal of Academic Language and Learning* 8 (3): A34–A48.

Maton, K. 2014b. *Knowledge & Knowers: Towards a Realist Sociology of Education*. London: Routledge.

Maton, K., and R. Moore. 2010. *Social Realism, Knowledge and the Sociology of Education: Coalitions of the Mind*. London: Continuum.

Mouton, M., and E. Archer. 2018. “Legitimation Code Theory to Facilitate Transition From High School to First-Year Biology.” *Journal of Biological Education*. Advance Online Publication 53 (1): 2–20. doi:10.1080/00219266.2017.1420681.

Nokes-Malach, T. J., and J. P. Mestre. 2013. “Toward a Model of Transfer as Sense-Making.” *Educational Psychologist* 48 (3): 184–207. doi:10.1080/00461520.2013.807556.

Podolefsky, N. S., and N. D. Finkelstein. 2006. “Use of Analogy in Learning Physics: The Role of Representations.” *Physical Review Special Topics – Physics Education Research* 2: 020101. doi:10.1103/PhysRevSTPER.2.020101.

Podolefsky, N. S., and N. D. Finkelstein. 2007a. “Analogical Scaffolding and the Learning of Abstract Ideas in Physics: An Example From Electromagnetic Waves.” *Physical Review Special Topics – Physics Education Research* 3 (1): 010109. doi:10.1103/PhysRevSTPER.3.010109.

Podolefsky, N. S., and N. D. Finkelstein. 2007b. “Analogical Scaffolding and the Learning of Abstract Ideas in Physics: Empirical Studies.” *Physical Review Special Topics – Physics Education Research* 3 (2): 020104. doi:10.1103/PhysRevSTPER.3.020104.

Rootman-le Grange, I., and M. A. L. Blackie. 2018. “Assessing Assessment: In Pursuit of Meaningful Learning.” *Chemistry Education Research and Practice* 19 (2): 484–490. doi:10.1103/PhysRevSTPER.3.020104.

Schoenfeld, A. H. 1999. “Looking Toward the 21st Century: Challenges of Educational Theory and Practice.” *Educational Researcher* 28 (7): 4–14. doi:10.3102/0013189X028007004.

Shay, S. 2008. “Beyond Social Constructivist Perspectives on Assessment: The Centring of Knowledge.” *Teaching in Higher Education* 13 (5): 595–605. doi:10.1080/13562510802334970.

Young, H. D., and R. A. Freedman. 2016. *Sears and Zemansky's University Physics with Modern Physics*. 14th ed. London: Pearson Education.