3 Improving assessments in introductory Physics courses

Diving into Semantics

Christine M. Steenkamp and Ilse Rootman-le Grange

Introduction

The discipline of Physics constitutes a highly 'hierarchical knowledge structure' (Bernstein & Solomon 1999) in which interrelated core concepts build on one another. Each core concept is applicable to a vast range of real-world contexts, though understanding and modelling of most real-world problems require the combination of several concepts. The ability to transfer knowledge to different contexts is critically important in Physics education where students are required to not only master the hierarchical knowledge structure of Physics but also apply the core concepts to different parts of the curriculum, to real-life scenarios and in future unfamiliar work environments (Laverty *et al.* 2016).

In introductory Physics modules, the priority is not only the success of the students in the module but also their degree of preparation for subsequent study. The challenge is to cultivate learning that facilitates the ability to 'transfer knowledge across contexts and build knowledge over time' (Maton 2009: 45). The concept of learning that facilitates transfer has been incorporated into the term 'cumulative learning,' highlighting that transfer is essential for students to build new knowledge on previously acquired knowledge (Maton 2009: 43). Cumulative learning also enables students to apply their knowledge in new contexts (Kilpert and Shay 2013).

The concept of 'transfer' originated in the field of Psychology (Nokes-Malach and Mestre 2013; Barnett and Ceci 2002). Transfer has been studied extensively in science disciplines (Lobato 2006) including Physics (for example, Finkelstein 2005). In a review of Physics education research, Docktor and Mestre (2014: 31) highlighted research on cognition as one of the prominent research directions, with 'learning and transfer' as a focus. In first-year Physics, cumulative learning means that students must know the core concepts in the curriculum and develop the ability to apply these in different contexts. Observation shows that this approach to learning is unfamiliar and difficult to first-year students (Walsh *et al.* 2007).

DOI: 10.4324/9781003055549-5

In an ongoing study, we are investigating the application of the Semantics dimension of Legitimation Code Theory (LCT) (Maton 2014b) as a framework for analyzing our educational practice in introductory Physics. The Semantics dimension was chosen as it offers a distinction between context-dependence and complexity, which are expected to be important factors in assessment questions. The study focuses on the analysis of questions in summative assessments of the first introductory Physics module in a three-year Bachelor of Science degree. The first reason for the focus on assessment questions is that 'assessment always acts as an intervention into student learning' (Bearman *et al.* 2016: 547) since assessment communicates in the most concrete way to students what they are expected to learn (Brown and Knight 1994; Laverty *et al.* 2016). The second reason is that it is very challenging to develop cumulative learning in a real academic environment where teaching styles vary, staff may change and it is difficult to enforce uniform teaching practice. However, the setting of summative assessments (tests and exams) is a regulated process involving all lecturers responsible for the module and the internal moderator, and it is possible to reach an agreement on how this process is to be executed and to implement changes. As assessment is a high-stakes activity for both students and teachers, we expect that a better understanding of what is assessed will influence both teaching and learning.

We consider this study important for the improvement of our own teaching and assessment practices. We as Physics teachers are comfortable with the language of Physics but often blind to the hurdles that exist for the students in acquiring not only the explicit knowledge but also the unwritten 'organizing principles of knowledge practices' (Georgiou *et al.* 2014: 255) that are typical of the discipline. In Physics, we have a language to discuss quantum mechanics but not to discuss the difficulty of questions in assessments (Johnston *et al.* 1998). This problem becomes apparent in the compulsory internal moderation process in our department. Each assessment paper is reviewed by a staff member who is not part of the teaching team for the module. However, this process typically does not result in significant changes to assessment questions, due to the lack of a framework for judgement (Fakcharoenphol *et al.* 2015; Beutel *et al.* 2017).

The Semantics dimension of Legitimation Code Theory

LCT provides a theoretical framework for the study of knowledge practices (Maton 2014b: 2). Two aspects of LCT make it accessible and attractive to natural scientists. Firstly, it makes knowledge a key object of study so that 'the nature of what is taught and learned' (Maton 2013: 9) and disciplinary expertise are given their rightful relevance. Secondly, it offers a suite of analytical tools suitable for empirical research into social practices (Maton 2014b; Maton and Moore 2010).

The Semantics dimension is one set of analytical tools in LCT. These tools foreground knowledge practices (including words, symbols, concepts, images and any other form of expression) and involve a distinction between context-dependence (the degree to which meaning is related to context) and complexity (the degree to which meanings are condensed into a practice) (Maton 2014a; Maton and Doran 2017). In Semantics, variation in context-dependence is termed *semantic gravity* and variation in complexity is termed *semantic density*. In the coding of semantic gravity, it is conventional to code knowledge practices that are strongly embedded in, for example, everyday experience as stronger semantic gravity and more general or abstract forms of knowledge as weaker semantic gravity. The strength of semantic density is coded as stronger as more meanings are condensed into knowledge practices. The strengths of semantic gravity and semantic density may each vary along a continuum, and strengths are always relative, never absolute. It is conventional to represent these concepts on the *semantic plane*, with semantic density and semantic gravity on the two axes (Figure 3.1) Four principal modalities are identified: *rhizomatic codes* (SG–, SD+), *prosaic codes* (SG+, SD–), *rarefied codes* (SG–, SD–) and *worldly codes* (SG+, SD+).

Semantic gravity is a measure of context-dependence that is directly connected to cumulative learning and transfer. Maton (2009) argues that exposing students to knowledge practices with a wide range of semantic gravity strengths is a key condition for fostering cumulative learning. Not only should students be exposed to knowledge practices with weaker semantic gravity, but the transition between context-independent principles and

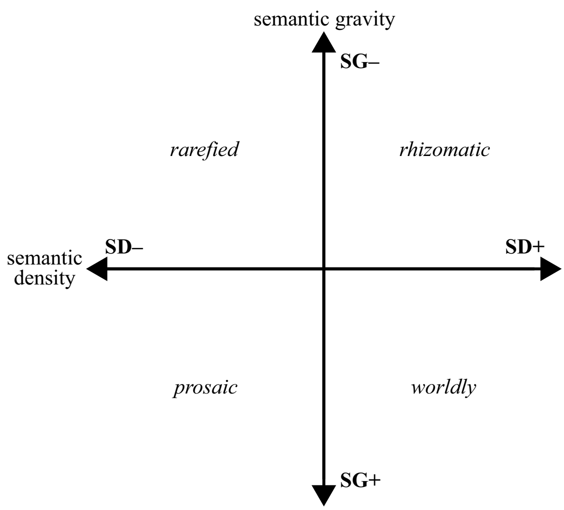


Figure 3.1 The semantic plane (Maton 2014a, 2014b: 131).

context-dependent examples, and vice versa (which are termed 'waves') should be modelled in teaching (Maton 2013). A number of papers (Kilpert and Shay 2013; Maton 2013; Conana *et al.* 2016) find that in their data, semantic gravity and semantic density are related in an inverse way: stronger semantic gravity (context-dependence) is associated with weaker semantic density (low complexity) and weaker semantic gravity (context-independence) is associated with stronger semantic density (high degree of condensation of knowledge in terms or other representations). However, Maton (2013) emphasizes that the two strengths can and do change independently so that one can encounter context-dependent but complex meanings and context-independent but simpler meanings.

An early application of LCT to assessment was done by Shay (2008) who concluded that LCT is a useful framework for conceptualizing the relation between knowledge and the criteria used in assessments. It has been concluded in several papers that cumulative learning is promoted by assessing over appropriate ranges of semantic gravity (Maton 2013; Kilpert and Shay 2013). Recently, the Semantics dimension was used to critique first-year Chemistry exam questions (Rootman-le Grange and Blackie 2018) and in our earlier work to analyze Physics assessments (Steenkamp *et al.* 2021). Semantics has also been applied to Physics education research by Georgiou *et al.* (2014) in analysis of students' responses to a thermal Physics question and by Conana *et al.* (2016) in a study of students' methods of problem-solving on the topic of mechanics in first-year Physics. In both papers, answers of students were analyzed using the wording of the question as context for exploring semantic gravity.

In our study, we apply semantic density and semantic gravity independently to analyze questions in Physics assessments. We agree that, in general, material with weaker semantic gravity tends to be expressed with stronger semantic density. However, in the context of our study objective to characterize assessment questions and answers, where a large variety of question types are included, we propose that there may be exceptions to this trend. Our approach of coding semantic gravity and semantic density independently is related to a study of Chemistry (Blackie 2014; Rootman-le Grange and Blackie 2018) but has to our knowledge not been applied before to question papers in Physics.

Background to the present study

The module on which the present study focuses is the first of two calculus-based introductory Physics modules (called Ph101 and Ph102, respectively, for the purpose of this study) that constitute the first year of a three-year Bachelor of Science programme. The modules are compulsory for students in experimental and theoretical Physics (12–20% of the class), Mathematics, Applied Mathematics, Computer Science, Earth Sciences, Geoinformatics, Chemistry, Polymer Science and are electives for students in various biological programmes. The Ph101 curriculum introduces vector algebra and basic calculus and covers Newtonian mechanics, gravity, fluid mechanics and thermodynamics. In Ph 102 electricity, magnetism and special relativity are the main topics. 'Sears and Zemansky's University Physics' authored by Young and Freedman (2016) is the prescribed textbook. The main summative assessments in Ph101 are two tests (tests 1 and 2) and two exam opportunities for each module (exams 1 and 2).

We have previously published an analysis of test and exam papers using semantic gravity of both Ph101 and Ph102 for 2012–16 (Steenkamp *et al.* 2021). This study showed significant variation in the semantic gravity ranges that were assessed, as well as a correlation between semantic gravity and students' average marks. We concluded that the range of semantic gravity in assessments should be controlled to ensure consistent and fair assessments. The second key finding was that questions from the weakest semantic gravity category, that test knowledge of core concepts, were underrepresented in the question papers. Therefore, our assessments did not communicate the importance of the core concepts to students. This analysis led to an informed internal moderation process of new test and exam papers (referred to as the intervention) during the 2017 academic year (Steenkamp *et al.* 2021). The intervention was aimed to change our assessments so that students will be assessed consistently over appropriate ranges of semantic gravity and to emphasize the importance of core concepts, by their direct assessment. Evaluation of the intervention resulted in the hypothesis that increased focus on and assessment of core concepts lead to an improvement in the students' ability to transfer their knowledge to real-life problems on the stronger semantic gravity end of the scale. The impact of semantic density was not considered during this original study.

Aims

This chapter reports on the analysis of semantic density in the Ph101 assessments of 2016 and 2017 (before and after the intervention). The analysis was motivated by the question of whether the manipulation of semantic gravity during the intervention also influenced the semantic density of assessment questions and, if so, how this influenced student performance. The semantic density analysis was done after the papers had been finalized and written. The overarching aim of this study is to determine what we can learn by using Semantics to analyze assessment questions in first-year Physics modules and whether such analysis is useful for making changes in educational practices.

By combining the current semantic density analysis with previously produced semantic gravity analysis data, we aimed to answer the following questions, using the intervention during 2017 as a test case:

- What does analysis of the context-dependence and complexity of assessment questions reveal about what we really require of students?
- Is there a correlation between the semantic density and semantic gravity strengths of questions or answers and students' ability to answer the questions (marks)?
- Did the intervention based on semantic gravity also unintentionally change the semantic density of the assessments?
- Would semantic density analysis support or falsify the hypothesis that the increased focus on core concepts during the intervention improved students' ability to transfer their knowledge to problems with stronger semantic gravity?

The current study is focused on the assessments of Ph101 as the intervention had the largest effect on this module. On the longer time scale, we aim to continue using Semantics for self-critique of our assessments in order to facilitate informed change to teaching and assessment practices.

Methodology

LCT concepts are highly abstract and general, in order to be applicable to a wide range of different practices. Thus, for each study one needs a 'translation device' (Maton 2014b; Maton and Chen 2016) that shows how a concept is realized in the particular data being analyzed. The development of a 'translation device' for enacting semantic density is discussed below. The semantic gravity analysis that we used has been published before, and thus the translation device is only discussed briefly (Steenkamp *et al.* 2021). Both devices were developed in order to be useful for all types of questions in our assessments.

The translation device for enacting semantic density, as shown in Table 3.1, considers the different representations of knowledge that are generally used in Physics to condense meaning. The use of representations is an active field in Physics education research (Van Heuvelen 1991; Fredlund *et al.* 2014; Docktor and Mestre 2014). In their paper pioneering the application of Semantics in Physics, Conana *et al.* (2016) also referred to different types of representations in their analysis of students' problem-solving approaches. The development of our translation device was guided by recent work by Maton and Doran (2017) on English discourse, which suggests that the more meanings are condensed within a representation, the stronger the semantic density. The typical representations used in Physics include verbal descriptions complemented by sketches and diagrams, more specialized vector diagrams or graphs and mathematical representations.

From weaker towards stronger semantic density, our translation device differentiates: verbal descriptions and images providing no technical information (SD−−), descriptions and images that does provide technical

Table 3.1 Translation device for semantic density

Category	Criteria
SD++	A mathematical representation that has been manipulated and condensed to construct a mathematical model for the problem.
SD+	A mathematical representation expressing relations between key concepts.
SD0	A graph conveying relationships between key concepts or a vector diagram conveying the relationships between the vector quantities in the problem.
SD–	Verbal representation of the problem using the key technical terms or symbols and/or a sketch or diagram conveying technical details.
SD––	Verbal representation of the problem in everyday words and/or a photo or sketch conveying the important features of the problem visually but using no technical terms and conveying no technical details.

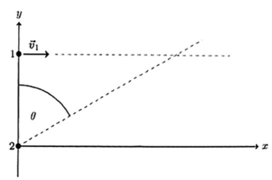
information that links them to core concepts (SD–) and graphs or vector diagrams (SD0) that provide information regarding relations of concepts that is more detailed than can be given in verbal descriptions. Mathematical representations consist of symbols representing physical quantities and occasionally contain numerical values. We consider a mathematical representation that expresses the relation between two or more key concepts to be typically of higher semantic density (SD+) than a graph of the same relation (SD0) as the mathematical expression usually contains additional meaning. Furthermore, the 'SD+' category simply requires core concepts and their basic relations to have been written down as mathematical expressions, while in the 'SD++' category mathematical expressions are manipulated to construct a mathematical model for a specific problem. The model then contains more meaning than the collection of expressions typical of the 'SD+' category, because relations that were not apparent in the 'SD+' category are now made explicit and typically the relation between parameters becomes more complex.

In the coding of semantic density, we disregard the numerical step that is sometimes required in assessments – namely, to substitute numbers into the mathematical expression and subsequently perform numerical calculations in order to obtain a numerical answer. In our assessments, this step usually counts very little in terms of mark allocation. Table 3.2 contains some examples of our coding of the various assessment questions, to further clarify the application of the translation device.

In the semantic density analysis, the assessment questions and model answers were coded independently. In compound questions the semantic density of each numbered sub-section was coded separately. When a question or an answer included elements of different semantic density categories,

Table 3.2 Examples of the coding of assessment questions representing different categories of semantic density and semantic gravity, along with an explanation to justify the allocated coding

Question	Model answer	Explanation of coding
Explain briefly what $\hat{j}$ represents in the context of vectors. Would it be correct to say that $\hat{j}$ is the same as 1? Explain your answer.	$\hat{j}$ represents the unit vector in the positive y direction. $\hat{j}$ is a vector with a direction and is therefore not the same as the scalar 1.	The semantic gravity is coded SG−− as there is no reference to any empirical example. The semantic density is relatively weak (SD−) for both question and answer as both are verbal descriptions including symbols with technical meaning.
A bat strikes a 0.145 kg cricket ball. Just before the impact, the ball is travelling horizontally to the right at 60.0 m/s. After being struck by the bat, the ball travels to the left at an angle of 35 degrees above the horizontal with a speed of 60.0 m/s. the ball and bat are in contact for 1.85 ms. Determine the horizontal and vertical components of the average force on the ball.	Write the momentum of the ball before and after being hit as vectors: $\vec{p}_1 = (60)\hat{i}$ $\vec{p}_2 = -(60)\cos(35\,degrees)\hat{i} + (60)\sin(35\,degrees)\hat{j}$ Change in momentum is related to average force via the impulse. $\vec{J} = \vec{p}_2 - \vec{p}_1 = \vec{F}\Delta t$ $\vec{F} = \frac{\vec{p}_2 - \vec{p}_1}{\Delta t}$ Substitute the numbers and calculate final vector.	The semantic gravity is relatively weak (SG−) as one core concept – namely, impulse – is applied to a simplified scenario closely associated with that concept. The semantic density of the question is coded as SD−− as it is a verbal description in everyday words. The semantic density of the answer is stronger (SD+) since the expected answer requires the use of mathematical expressions.

At a time $t = 0$ s a particle 1 has the position $r_1 = 30\,\hat{j}$ m. It always moves at a constant velocity $v_1 = 3.0\,\hat{i}$ m/s. Another particle, particle 2, has zero velocity at time $t = 0$ s, and it has a constant acceleration $\vec{a}$, with magnitude $a = 0.40$ m/s². Particle 2 is at the origin at $t = 0$ s. There is no gravity. Calculate at which angle θ the acceleration vector $\vec{a}$ must point with respect to the positive y-axis so that the two particles v collide. 	The student must apply this argument: we need to compute the trajectories of both particles. They will collide if the x and y components of the two trajectories are identical at the same time. Particle 1 has a constant velocity: $\vec{r_1}(t) = (3.0)t\,\hat{i} + (30)\,\hat{j}$ Particle 2 starts at origin with zero velocity but constant acceleration. The magnitude of its displacement is $r_2(t) = \frac{1}{2}(0.4)t^2\,\hat{i} + (30)\,\hat{j}$ Written as vector $\vec{r_2}(t) = \frac{1}{2}(0.4)\sin\theta t^2\,\hat{i} + \frac{1}{2}(0.4)\cos\theta t^2\,\hat{j}$ Set x and y components equal at the time of collision: x: $(3.0)t = \frac{1}{2}(0.4)\sin\theta t^2$ y: $30 = \frac{1}{2}(0.4)\cos\theta t^2$ Combine these relations and solve simultaneously (a fair amount of work still) to obtain $\theta = 60$ degrees.	The semantic gravity is relatively strong (SG++) as this problem requires the student to construct a self-defined logical argument, involving the translation of everyday knowledge of a collision into a mathematical condition. The question contains a verbal description with technical terms (SD−) and a graph (SD0). The answer has stronger semantic density (SD++) as it requires significant manipulation and condensation of mathematical relations.
Water has the important property that water has the highest density at 4 degrees Celsius. Water at higher or lower temperatures, and ice, have lower densities. Explain why this property is important to prevent large water bodies (like lakes) from freezing solid down to the bottom.	The answer requires a verbal explanation, including discussion that this property of water prevents convection in water at temperatures lower than 4 degrees Celsius, thus freezing from the top and that a surface layer of ice acts as thermal insulator.	The semantic gravity is relatively strong (SG++) as the student must combine the given information and core principles, combined with everyday knowledge, to a specific 'real-world' scenario. The semantic density of both the question and answer is SD−, as both consist of verbal representations using technical terms.

we reported the strongest semantic density. The motivation for this decision is that every question or answer will include text of weaker semantic density, and it is of little meaning to code that if elements of stronger semantic density are present.

The translation device for semantic gravity has been published before, and a more detailed discussion is given in the original publication (Steenkamp *et al.* 2021). The translation device is reproduced in Table 3.3. For this translation device, four categories were defined, labelled: SG−−, SG−, SG+ and SG++. The weakest category, 'SG−−' , is associated with formulating a core concept in the general context-independent form, 'SG−' is the application of a concept to a simplified and idealized scenario, 'SG+' is the application to a well-defined empirical scenario and 'SG++' represents the application of core concepts to a real-world problem, where self-defined assumptions and application of everyday knowledge is required in addition to the concepts. In the case of semantic gravity, the question with its model answer is coded as a unit, contrary to the independent coding of question and answer in semantic density. Most assessment questions include aspects of more than one semantic gravity category. For the purpose of combining our semantic gravity and semantic density analyses, we reported the strongest semantic gravity categories found in those questions. Examples of how the translation device was applied to our data are presented in Table 3.2.

The average marks of the cohort of students for individual questions, or sub-questions, were used as a measure of the students' ability to master the semantic gravity and semantic density present in the question and required for a successful answer. In order to compare marks before (2016) and after the intervention (2017) the group of first-time enrolled students were selected, and the marks of students who were repeating the module were not considered. The data also excluded the small number of students who deregistered or changed to a different programme during the academic year.

Table 3.3 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.

From Steenkamp *et al.* (2021), reprinted by permission of the publisher (Taylor & Francis Ltd, http://www.tandfonline.com)

For a detailed comparison, the marks to the first exam of Ph101 were analyzed on the level of sub-questions. This was done for 50% of the cohort, selecting every second name on an alphabetical list as a representative sample.

Results and discussion

Figure 3.2 shows the semantic density analysis of the assessments of 2016 and 2017. The graphs show the percentage of marks allocated to either questions (a) or answers (b) for each of the four semantic density categories. We will call these the semantic density 'maps' of the assessment questions or answers.

It is observed that the questions generally show weaker semantic density than the model answers. The semantic density of the questions include a few questions in the 'SD−−' category (non-technical verbal descriptions), a large majority of questions in the 'SD−' category (verbal descriptions involving technical terms), the use of graphs and vector diagrams in the 'SD0' category and of mathematical expressions in the 'SD+' and 'SD++' categories. The model answers generally exclude non-technical verbal descriptions (SD−−). Technical verbal descriptions (SD−) comprise on average 10%–20% of the answers. Graphs and diagrams (SD0) form a small fraction of answers, writing down basic mathematical expressions without significant manipulation (SD+) makes up 20–30% of answers. More than 50% of answers require the construction of mathematical models by manipulating and condensing the basic relations to derive additional relations (SD++).

It is evident that the intervention caused observable changes in the semantic density maps. In both the questions and the answers, the semantic density weakened on average. In both questions and answers, the use of verbal descriptions using technical terms (SD−) increased while the use of mathematical expressions (SD+) and models (SD++) decreased during 2017 in comparison to 2016. This reflects a larger focus on core concepts and a

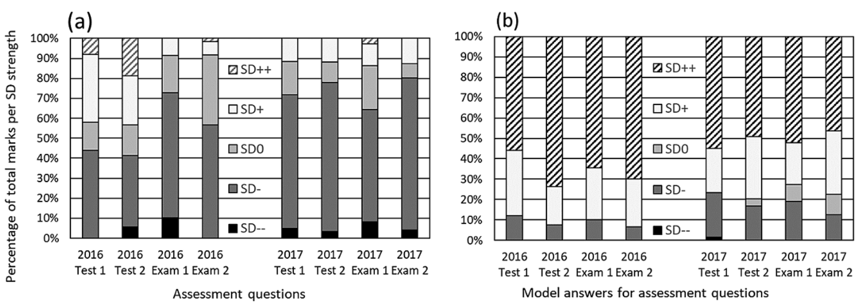


Figure 3.2 Semantic density maps of the Ph101 assessments of 2016 and 2017 representing the percentage of marks allocated to (a) questions and (b) model answers of different semantic density categories.

realization that assessment of conceptual understanding should include verbal explanations. Compared to the semantic gravity maps of the same assessments (see Figure 3.4(a) of Steenkamp *et al.* 2021), the change in the semantic density map is less pronounced than the change in the semantic gravity map. This is probably because the semantic gravity map was modified on purpose, whereas the change in semantic density was unintentional.

The use of mathematical expressions in questions during 2016 and 2017 differs in an interesting way. In both years, some questions contain mathematics that form an integral part of the question, but questions also occur where non-essential mathematical expressions are given. These questions can be answered without the mathematical expression, but the expression serves as 'scaffolding' (Dawkins *et al.* 2017). Considering these questions, the expressions given during 2016 were often the expression for the final answer that must be derived or proven (coded as SD++). This communicates an emphasis on reaching a certain answer. During 2017, the given expressions were usually representing one of the core concepts that serve as a starting point for answering the question (coded as SD+), communicating an emphasis on core concepts. This may be the result of the lecturers' increased awareness of core concepts.

Graphs and vector diagrams (SD0) were used in questions in both years but were present in answers only during 2017. In these 2017 questions, conceptual understanding was tested by requiring students to sketch a graph or a vector diagram.

We found it useful to characterize assessment question-answer pairs by plotting these as points on a graph that has the semantic density of the question on the horizontal axis and the semantic density of the answer on the vertical axis (Figure 3.3). The areas of the circles in Figure 3.3 represent the number of question-answer pairs in each category. What this analysis revealed is that most of the data points are above the upwards sloping diagonal, meaning that the model answers generally have stronger semantic density than their associated questions. However, there are questions in each test positioned below the

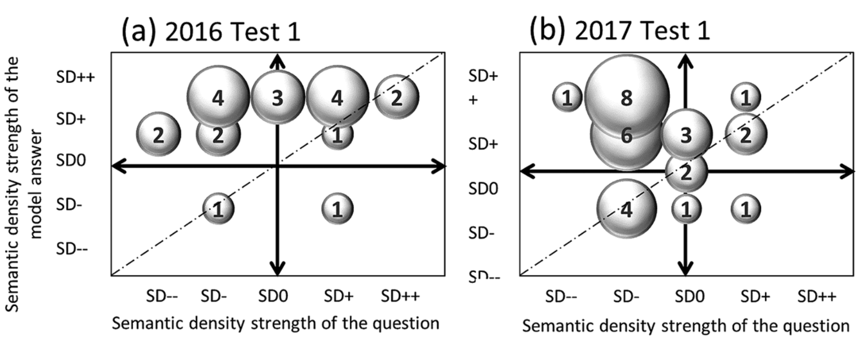


Figure 3.3 Number of question-answer pairs plotted on a two-dimensional graph representing the semantic density of answers versus questions, for the main tests of 2016 (a) and 2017 (b), respectively.

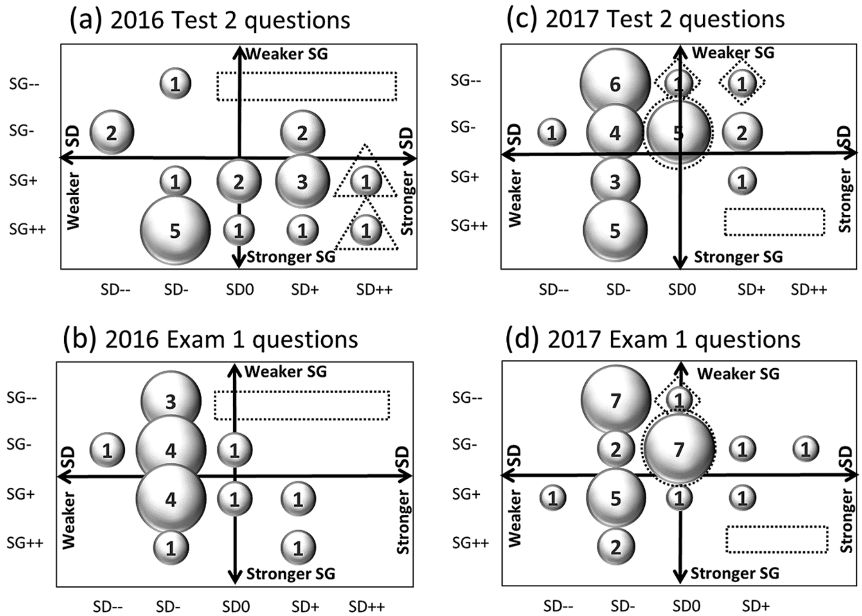


Figure 3.4 The questions of the class tests of 2016 (a) and 2017 (c) and first exams of 2016 (b) and 2017 (d) of Ph101, plotted on the semantic plane.

diagonal, where a verbal description (of weaker semantic density) was required as answer for a question containing a diagram or mathematics (of stronger semantic density). In the 2016 test, the model answers were mostly in the form of mathematical expressions (SD+ and SD++ on the vertical axis), whereas a wider distribution of answer types is seen in the 2017 test. We conclude that the purposeful changes to the semantic gravity maps of the question papers during 2017 did influence the semantic density of both the questions and required answers.

The semantic plane offers a visualization of the relation between semantic gravity and semantic density. Figure 3.4 shows the semantic planes for the questions from four analyzed assessments, while Figure 3.5 shows the semantic planes for the model answers of these assessments. In both figures, the areas of the circles and numbers inside represent the number of question-answer pairs in each category. Dotted lines are guides for the eye and are referred to in the discussion.

Figure 3.4 shows that the questions cover the full range of semantic gravity and semantic density, with the SD– category the most prominent in terms of semantic density. In the 2016 papers (also typical for the other 2016 assessment not shown here), there is a trend that a stronger semantic density is associated with a stronger semantic gravity and vice versa, seen by the dominant grouping of questions in the rarefied and worldly codes, in comparison to the rhizomatic and prosaic codes. Furthermore, questions coded 'SG––,' that directly assess core concepts, are asked as technical verbal

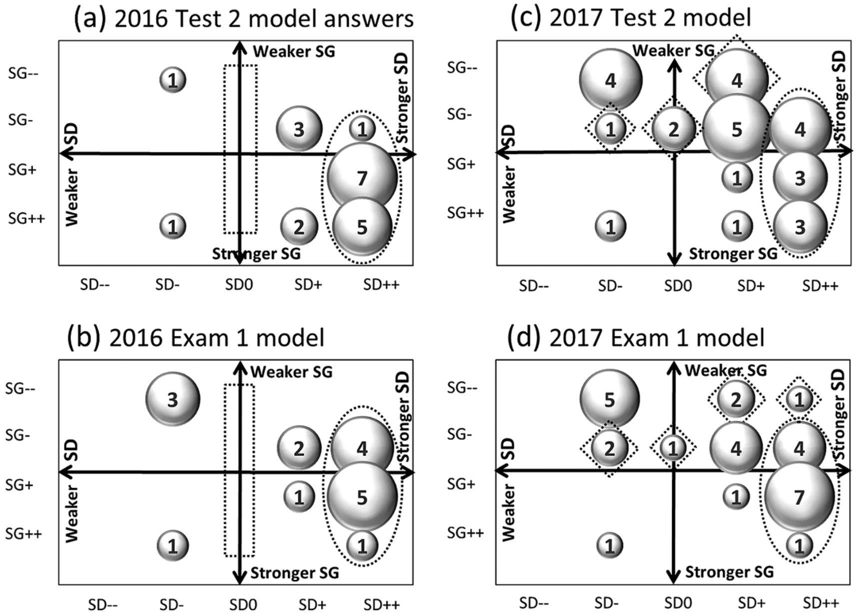


Figure 3.5 The model answers of the class tests of 2016 (a) and 2017 (c) and first exams of 2016 (b) and 2017 (d) of Ph101, plotted on the semantic plane.

descriptions (SD–) only, and questions of semantic gravity SG–– that have stronger semantic densities (indicated by dotted rectangles in graphs (a) and (b)) are typically lacking. Questions in worldly codes (dotted triangles) with both stronger semantic gravity (SG+ and SG++) and stronger semantic density (SD++) were frequently asked before the intervention. These trends differ from the association of weaker semantic gravity with stronger semantic density used in other studies (Conana *et al.* 2016). The reason for this is that before the intervention, typically questions that required applications to real-life scenarios (SG+ or SG++) had to be solved by constructing a mathematical model, and some of that mathematics was given in the question as scaffolding (SD+ or SD++), whereas questions on core concepts (SG–– or SG–) were mostly assessed using verbal descriptions (SD–– or SD–).

In 2017, this trend was absent. The semantic gravity category 'SG––,' where core concepts are assessed directly, was assessed in more diverse ways, using questions with stronger semantic density (dotted diamond shapes). The number of questions where a graph or vector diagram is given in the question and assessed within an intermediate semantic gravity range (dotted circles) was increased. This resulted in a distribution of questions between the rarefied, rhizomatic and prosaic codes.

The lack of questions in the worldly code representing stronger semantic gravity and stronger semantic density – rectangles in graphs (c) and (d)) in the 2017 assessments – may be flagged. At first impression, it may be asked whether the 'hardest' questions have been omitted from the 2017 assessments. However, closer analysis revealed that in the two SD++ questions in the 2016 test, the mathematical expression of the final answer is given, and the student is asked to prove that this is true. The same Physics has been assessed in the 2017 test without giving the final answer as part of the question, thus weakening the semantic density of the question without changing the Physics that is assessed.

A meaningful use of the SD++ question category would be to give a mathematical model of a real-life scenario and ask students to interpret the model and draw conclusions. This example illustrates how plotting the questions on the semantic plane can be useful for critical self-evaluation of a question paper by giving an overview of the types of questions, but that not all questions in a particular SG and SD category are equal in terms of difficulty.

Figure 3.5 shows the coding of the model answers on the semantic plane. A clear difference between 2016 and 2017 assessments can be seen. In the 2016 assessments, most of the answers are clustered in the worldly code (dotted oval in Figure 3.5(a)) representative of stronger semantic density (SD++) and intermediate to stronger semantic gravity categories (SG–, SG+, SG++). Furthermore, students were not asked to draw graphs or vector diagrams (SD0) as answers (dotted rectangle). Answers in the 'SG––' category are only in the form of technical verbal descriptions (SD–). It thus shows that the intervention caused us to set questions that require answers over a wider variety of semantic gravity and semantic density categories. New types of questions (indicated by dotted diamond shapes) include answers on core concepts (SG––) with stronger semantic density (SD+) and applications of core concepts (SG–) that require a graph (SD0) or verbal description (SD–). Answers requiring the construction of a mathematical model (SD++) remain important, as indicated by the dotted ovals in Figure 3.5(c) and (d).

The results confirm changes in the semantic density, as well as semantic gravity, due to the intervention. In 2017, the questions and answers are distributed more evenly over a wider range of semantic gravity and semantic density categories. An increased focus on weaker semantic gravity (SG––) is also observed. In terms of semantic density, technical verbal representations as questions and mathematical expressions as answers remain dominant, corresponding to the semantic density maps in Figure 3.2.

An important question remains to be answered: do student marks correlate with question-answer pairs of different semantic gravity and semantic density codes? To investigate this, we performed a detailed analysis of the average marks of the first exams of 2016 and 2017 on the level of subquestions in order to calculate the average mark for questions of a particular semantic gravity category and semantic density category. We selected exam 1 since it is the most important assessment. Results from the 2016 and 2017 papers were combined.

We have found that the most useful way to study the effect on marks is to visualize the dependence of marks on semantic gravity and semantic density as in Figure 3.6. Figure 3.6(a) and (c) show the variation of marks over the different semantic density categories of the questions and answers,

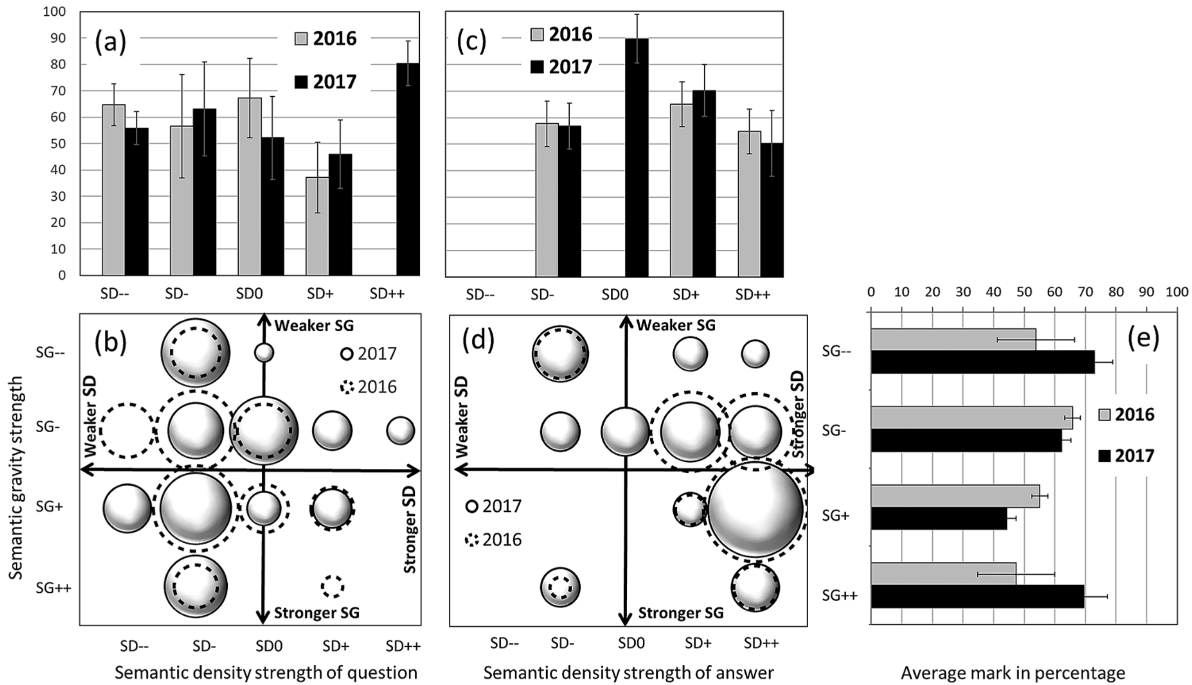


Figure 3.6 The average marks plotted versus the semantic density of the question (a) and answer (c) and the semantic gravity (e), respectively, shown with the weights of the different categories of questions (b) and answers (d) plotted on the semantic plane.

respectively, whereas Figure 3.6(e) shows the variation of marks over the different semantic gravity categories for question-answer-pairs combined. The error bar lengths in these graphs agree to the standard deviations of the marks. To facilitate interpretation of these graphs, they are displayed with plots of the weights of the question-and-answer categories on the semantic plane – Figure 3.6(b) and (d). In these plots, the diameter of each circle represents the number of marks allocated to the category as percentage of the total of the exam paper. It means that, for example, all the questions represented in the vertical SD+ column of graph (b) contribute to the marks represented by the SD+ bar in graph (a), and all the questions represented in the horizontal SG– row of graph (b) contribute to the marks in the SG– bar in graph (e).

The variation of marks over the semantic gravity categories (Figure 3.6(e)) shows different trends during 2016 and 2017. During 2016, the marks were highest for intermediate semantic gravity categories (SG– and SG+) and students did not perform as well as expected in the weakest category (SG––). This means that students struggled to identify the correct core concept or to formulate or explain the concept correctly. Marks are also lower, as expected, when the semantic gravity is stronger (SG++), and a significant degree of transfer of knowledge to a real-life scenario is required. During 2017, this trend became inverted so that the weakest category (SG––) correlated with the highest average mark. This confirms that the efforts to encourage students to know the core concepts well have been successful. The surprising result was that the marks in the strongest category (SG++) also increased significantly. The question was asked whether these differences could be caused by unintended changes in semantic density. This question can be investigated using Figure 3.6, as the plot on the semantic plane links the graph of marks versus semantic gravity to the graph of marks versus the semantic density.

In the weakest semantic gravity category (SG––) in Figure 3.6(b), the questions were of approximately the same semantic density (mostly SD–) during 2016 and 2017. The answers (Figure 3.6(d)) were extended to stronger semantic density (SD+ and SD++) during 2017. It means that the 2017 students performed better in questions testing core concepts, although answers of stronger semantic density were expected in this category. In the strongest semantic gravity category (SG++), the semantic density of questions (Figure 3.6(b)) and answers (Figure 3.6(d)) during 2016 and 2017 were similar with a slightly higher weight given to weaker semantic density questions and answers in 2017, meaning that the improved performance of students in the SG++ category during 2017 cannot be the result of a significant change in semantic density. This supports the hypothesis that the higher marks in the weaker and stronger semantic gravity categories (SG–– and SG++) are indeed linked: the increased awareness of what the core concepts are and being able to formulate them correctly enables students to transfer their knowledge to real-life scenarios.

In SG– and SG+ categories, the marks are lower during 2017 than during 2016, but the decreases are relatively small (4% and 11% lower in the

SG– and SG+ categories, respectively). In the SG– category, the semantic density of the questions was extended towards stronger semantic density and the answers towards weaker semantic density during 2017. In the SG+ category, the questions were extended towards weaker semantic density and the semantic density of the answers is unchanged from 2016 to 2017. This means that the marks of the SG– and SG+ categories show a similar decrease from 2016 to 2017 in spite of opposing changes to the semantic density from 2016 to 2017.

The general conclusion is therefore that semantic density is not clearly correlated with the changes in marks from 2016 to 2017 in either of the four semantic gravity categories. Considering the variation of marks with the semantic density of the question (Figure 3.6(a)) a significant difference between the 2016 and 2017 marks (difference larger than the error bar representing the standard deviation) exists only in the SD–– category. Considering the data of both 2016 and 2017 together (and ignoring the SD++ bar of 2017 as it represents a single low-weight question only), the general trend is a decrease in marks as the semantic density of the question is increased. In this trend, the semantic gravity should not play a large role as the mapping of the questions on the semantic plane (Figure 3.6(b)) is approximately symmetric around the horizontal axis so that the trend must be caused by the increase in semantic density. The variation of marks with the semantic density of the answers (Figure 3.6(c)) shows a pronounced decrease of marks from SD0 to SD++. We consider this to be caused by both the combination of the increasing semantic gravity and increasing semantic density of the required answers.

We thus found evidence of an influence of semantic density on marks, but this effect is weaker than that of the semantic gravity. It thus seems that our students are more successful in working with different representations of knowledge than to transfer knowledge to unfamiliar contexts. The influence of increasing semantic density on marks is most pronounced in the SD0, SD+ and SD++ where the increasing semantic density is associated with graphs and symbolic mathematical representations. This correlates with observational evidence that many first-year students struggle to manipulate and interpret symbolic mathematical expressions and graphs.

The final question regarding the influence of semantic density on marks is whether it may be the difference between the semantic density of the question and the semantic density of the answer that plays a role, rather than any one of these separately. This is investigated by plotting the average marks versus the semantic density difference (SDD) in Figure 3.7, where the sign and number indicate the measure of strengthening (+) or weakening (–) of the semantic density from question to answer. The error bar length agrees to the standard deviation of the marks. For example, SDD+1 represents a question-answer pair for which the semantic density of the answer is one category stronger than that of the question. In this graph, the 2016 and 2017 marks show opposing trends. When ignoring the low number of data points in the SDD–2 and SDD–1 categories, the 2017 marks decrease

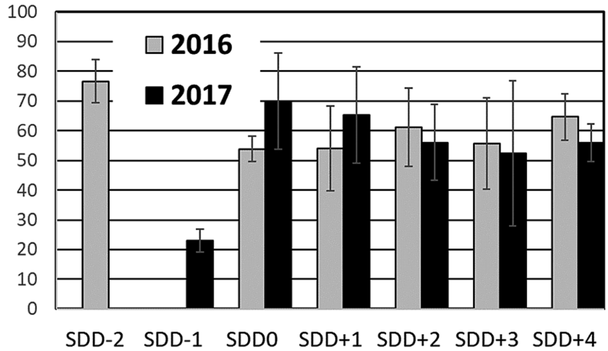


Figure 3.7 Average marks in the first exams of 2016 and 2017 as a function of the SDD.

towards larger SDDs, whereas the 2016 marks show an increasing trend over this range. We conclude that the SDD on its own cannot be a critical quantity.

The analysis including both semantic density and semantic gravity confirmed the conclusion previously proposed on the basis of semantic gravity only (Steenkamp *et al.* 2021). The intervention caused a significant increase in marks in both the weaker and the stronger ends of the semantic gravity scale used. We propose that the focus on core concepts in teaching and assessment caused by the intervention communicated the importance of core concepts clearly to students, encouraging them to study these. Improved familiarity with the core concepts and their correct formulation enabled students to transfer their knowledge to problems linked to real-life scenarios. As the core concepts in Physics are formulated to be highly context-independent (abstract), this result corresponds with the idea that significant exposure of learners to knowledge of weaker semantic gravity is not only supportive of but also a condition for cumulative learning (Maton 2009; Maton 2013).

Conclusion

We have used the Semantics dimension of LCT to evaluate the outcomes of an intervention in an introductory Physics module. The intervention was the result of a previous study, which employed semantic gravity to analyze the context-dependence of the assessment questions in historical papers and resulted in an internal moderation process during which the results were used to evaluate and modify new papers (Steenkamp *et al.* 2021). We conclude that the analysis of assessments using Semantics is a practical starting point for making changes in introductory Physics modules. We argue that assessment has a direct effect on learning as it is a high-stakes activity for students and lecturers, and it communicates the expected outcomes of learning concretely to students. Our study of question papers before and during the intervention (2016 and 2017) showed the value of both semantic gravity and semantic density in categorizing types of questions and model answers in assessments. Correlation of the question categories with students' average marks has confirmed that semantic gravity has the dominant effect on marks. We have observed a trend of decreasing average marks with increasing semantic density of either the question or the answer, but the effect is weaker than that of the semantic gravity.

The analysis including both semantic density and semantic gravity supported the hypothesis that an increased focus on core concepts (weaker semantic gravity) in teaching and assessment improved the ability of our students to transfer their knowledge to questions related to real-life scenarios (stronger semantic gravity). This is a step towards cumulative learning.

We conclude that the Semantics dimension of LCT is a valuable analytical tool to Physics lecturers and Physics education specialists. LCT acknowledges that disciplinary knowledge has a unique character and aims to study this in a systematic way, thus avoiding a 'knowledge-blind' approach to education (Maton 2013). We agree with the conclusions of Georgiou *et al.* (2014) that in Physics education the forms that knowledge takes play an important role in learning and that LCT, therefore, offers a productive approach to analyze educational practice and, in particular, evaluate changes in educational practice in the interest of informed decisions. LCT is, however, more than a tool, but a theoretical framework for the study of knowledge and useful to guide changes to many aspects of education.

Acknowledgements

The authors wish to thank Zhigang Yu, Tongji University, for helpful discussions about the semantic density of images. We wish to thank Margaret Blackie for valuable advice. Christine M. Steenkamp acknowledges the support from colleagues Kristian Müller-Nedebock, Gurthwin Bosman and Hannes Kriel.

References

Barnett, S. and Ceci, S.J. (2002) 'When and where do we apply what we learn? A taxonomy for far transfer,' *Psychological Bulletin* 128(4): 612–637.

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M. and Molloy, E. (2016) 'Support for assessment practice: Developing the assessment design decisions framework,' *Teaching in Higher Education* 21(5): 545–556.

Bernstein, B. and Solomon, J. (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.

Beutel, D., Adie, L. and Lloyd, M. (2017) 'Assessment moderation in an Australian context: Processes, practices, and challenges,' *Teaching in Higher Education* 22(1): 1–14.

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.

Brown, S. and Knight, P. (1994) *Assessing learners in higher education*, Kogan Page.

Conana, H., Marshall, D., & Case, J. M. (2016). Exploring pedagogical possibilities for transformative approaches to academic literacies in undergraduate Physics. *Critical Studies in Teaching and Learning*, *4*(2), 28–44.

Dawkins, H., Hedgeland, H. and Jordan, S. (2017) 'Impact of scaffolding and question structure on the gender gap,' *Physical Review Physics Education Research* 13(2): 020117.

Docktor, J.L., and Mestre, J.P. (2014) 'Synthesis of discipline-based education research in Physics,' *Physical Review Physics Education Research* 10(2): 020119.

Fakcharoenphol, W., Morphew, J.W. and Mestre, J.P. (2015) 'Judgements of physics problem difficulty among experts and novices,' *Physical Review Physics Education Research* 11(2): 020128.

Finkelstein, N. (2005) 'Learning physics in context: A study of student learning about electricity and magnetism,' *International Journal of Science Education* 27(10): 1187–1209.

Fredlund, T., Linder, C., Airey, J. and Linder, A. (2014) 'Unpacking physics representations: Towards an appreciation of disciplinary affordance,' *Physics Education Research* 10(2): 020129.

Georgiou, H., Maton, K. and Sharma, M. (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.

Johnston, I.D., Crawford, K. and Fletcher, P.R. (1998) 'Student difficulties in learning quantum mechanics,' *International Journal for Science Education* 20(4): 427–446.

Kilpert, L. and Shay, S. (2013) 'Kindling fires: Examining the potential for cumulative learning in a Journalism curriculum,' *Teaching in Higher Education* 18(1): 40–52.

Laverty, J.T., Underwood, S.M., Matz, R.L., Posey, L.A., Carmel, J.H., Caballero, M.D., Fata-Hartley, C.L., Ebert-May, D., Jardeleza, S.E. and Cooper, M.M. (2016) 'Characterizing college science assessments: The three-dimensional learning assessment protocol,' *PLoS One* 11(9): e0162333.

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.

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.

Maton, K. (2013) 'Making semantic waves: A key to cumulative knowledge building,' *Linguistics and Education* 24(1): 8–22.

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 and Knowers: Towards a Realist Sociology of Education*, Routledge.

Maton, K. and Chen, R. T-H. (2016) 'LCT in qualitative research: Creating a translation device for studying constructivist pedagogy,' in K. Maton, S. Hood and S. Shay (eds) *Knowledge-Building: Educational Studies in Legitimation Code Theory*. Routledge, 27–48.

Maton, K. and Doran, Y.J. (2017) 'Semantic density: A translation device for revealing complexity of knowledge practices in discourse, part 1—wording,' *Onomázein* (1): 46–76.

Maton, K. and Moore, R., eds. (2010) *Social realism, knowledge and the sociology of education: Coalitions of the mind*. Continuum.

Nokes-Malach T.J. and Mestre, J.P. (2013) 'Toward a model of transfer as sense-making,' *Educational Psychologist* 48(3): 184–207.

Rootman-le Grange, I. and Blackie, M.A.L. (2018) 'Assessing assessment: In pursuit of meaningful learning,' *Chemistry Education Research and Practice* 19(2): 484–490.

Shay, S. (2008) 'Beyond social constructivist perspectives on assessment: The centring of knowledge,' *Teaching in Higher Education* 13(5): 595–605.

Steenkamp, C. M., Rootman-le Grange, I., & Müller-Nedebock, K. K. (2021). Analysing assessments in introductory physics using semantic gravity: Refocussing on core concepts and context-dependence. *Teaching in Higher Education*, *26*(6), 871–886.

Van Heuvelen, A. (1991) 'Learning to think like a physicist: A review of research-based instructional strategies,' *American Journal of Physics* 59: 891–897.

Walsh, L.N., Howard, R.G. and Bowe, B. (2007) 'Phenomenographic study of students' problem solving approaches in Physics,' *Physics Education Research* 3(2): 020108.

Young, H.D. and Freedman, R.A. (2016) *Sears and Zemansky's University physics with modern physics* (14th ed.). Pearson Education.