

Joined up thinking? Evaluating the use of concept-mapping to develop complex system learning

Martyn Stewart*

Academic Enhancement Unit, Liverpool John Moores University, Kingsway House, Hatton Garden, Liverpool L3 2AJ, UK

In the physical and natural sciences, the complexity of natural systems and their interactions is becoming better understood. With increased emphasis on learning about complex systems, students will be encountering concepts that are dynamic, ill-structured and interconnected. Concept-mapping is a method considered particularly valuable for enabling learning in subject areas that are complex or ill-structured. Evaluations of concept-mapping tend to reflect their many applications. Many evaluations that try to measure enhancements in learning use test scores or grade-point averages as performance indicators, which provide little information on how cognitive processes have developed. In this study, a modification of the Biggs and Collis' Structure of the Observed Learning Outcomes (SOLO) taxonomy is applied to measure differences in the cognitive and structural complexity of learning outcomes between groups of students who participated in a concept-mapping activity with those from a cohort that did not. The evaluation demonstrated that the intervention was effective in enabling the great majority of students to achieve better connectivity in thinking, though improvements in overall performance were less significant. Almost all students perceived the intervention to be of value to their learning.

Keywords: concept-mapping; complex learning; relational learning; evaluation

Introduction

In the past decade there have been calls to reconsider the focus of higher education in the development of students' high-order thinking skills. Across many sciences, for example, the emergence of complex system 'theories' is shifting attention towards aspects of real-world system complexity that were previously ignored or oversimplified. The implications for how students learn this subject matter can be significant (Jacobson and Wilensky 2006). Elsewhere, developments to strengthen links between university research and undergraduate teaching have highlighted the intellectual gains for students of working with 'messy', real data and engaging first-hand with the processes of knowledge creation (Jenkins 2004; Jenkins, Healey, and Zetter 2007). In the context of personal development, it has been argued that universities need to be reformulated to better prepare students and society to deal with a world that is increasingly 'supercomplex' (Bowden and Marton 1998; Barnett 2000).

The educational importance of learning complex subject matter is realised, alongside the challenges this poses for student learning (Goldstone 2006; Hmelo-Silver and Azevedo 2006; Jacobson and Wilensky 2006). With increased emphasis on learning

*Email: m.stewart@ljmu.ac.uk

complex systems, students will be encountering subject concepts that are dynamic, multiplex, often poorly structured and highly interconnected, and will develop new abilities, mental models and attitudes. Spector (2006), Jacobson and Wilensky (2006) and Knight and Page (2007) suggest this will require a rethink in our methods for teaching and assessing learning, with different approaches to be called for at different stages of the continuum towards more 'expert' learning.

Concept-mapping is a constructivist teaching strategy considered valuable particularly for subject areas that are complex or ill-structured (Novak 1990; Clayton 2006; Spector 2006; Novak and Cañas 2007). The approach has been applied across education sectors in the past two decades for a range of purposes, including as a planning tool for teaching, as a learning activity, as an assessment method and as a research tool. Evaluations of the impact of the method vary in accordance with the different uses. When applied as a learning method, evaluations often consider the impact of the activity on students' overall performance reported as a test score (Fraser and Edwards 1985; Rooda 1994), but there are fewer studies that measure how the method develops thinking structure.

This study evaluates the impact of a concept-mapping intervention at Level 3 of a geological sciences course studying geotechnical engineering solutions to natural hazard events, such as landslides and floods. Through the real-world case studies, students are introduced to excellent examples of complex systems because the underlying causes and trigger mechanisms to the hazards are typically multiple, inter-related and highly site-dependent. A key learning outcome of the course is to develop students' understanding of the complexity of these dynamic systems, and concept-mapping exercises were introduced to actively engage students in exploring subject connections. This study examines whether the concept-mapping intervention resulted in more complex understandings expressed as learning outcomes in subsequent coursework.

Complex systems in science education

Jacobson and Wilensky (2006) summarise the implications of advances in complex system theory for education. The science of complex systems is emerging and, rather than a single unifying 'theory', encapsulates branches of developments within the mathematical, physical, natural and social sciences. In the context of this study, the interest is in natural systems that serve as good examples. Natural systems are typically dynamic and changing over time, are often held in states of equilibrium with other interdependent systems, and interactions can be unpredictable. Disturbing the 'web' of interconnections can have major implications as effects cascade across associated networks. A familiar example is that of climate change. An increase in global temperature will lead to thermal expansion of the oceans and melting of icecaps, which raises sea level. This in turn impacts upon associated systems, for example increasing coastal erosion and sedimentation patterns and disturbing local climate systems. These in turn can impact upon the stability of ecosystems, hydrological systems and so on.

Many of the core ideas associated with thinking about complex systems are cognitively challenging for students to learn (Hmelo-Silver and Azevedo 2006; Jacobson and Wilensky 2006) and might require students to go through processes of significant conceptual change, as occurs in the transition in thinking from novice to expert (Chi 1992, 2005).

The problems of teaching for complex understandings

Teaching strategies are highly significant in influencing how students derive their understanding of a subject, particularly in terms of its conceptual structure (Prosser, Trigwell, and Taylor 1994; Hay, Kinchin, and Lygo-Baker 2008; Kinchin, Chadha, and Kokotailo 2008). How a teaching session is constructed and delivered can vary greatly depending on wide range of factors such as lecturers' beliefs and knowledge about teaching, the epistemologies and traditions of the subject, the nature of the learning environment and even use of resources, such as learning technologies.

A major factor in determining a teaching design is the teacher's knowledge of pedagogic principles and capacity to apply theory to practice. Research into university academics' beliefs about teaching suggests the existence of a 'nested hierarchy' of conceptions that extend in sophistication from teacher- and content-orientations involving the simple transmission of syllabus content, to sophisticated student- and learning-centred orientations in which the aim is to develop and change students' conceptions through active learning (Prosser, Trigwell, and Taylor 1994; Kember 1997; Entwistle and Walker 2000). These differing intentions have been shown to relate logically to the teaching strategies deployed (Trigwell, Prosser, and Taylor 1994) with those teachers of the information-transmission orientation favouring the didactic lecture in which teaching quality is judged on criteria such as the accuracy, structure and speed of presentation, rather than the quality of students' learning. Where the intention is to develop higher-level learning outcomes associated with learning complex systems, research has demonstrated that active learning is particularly important (Jonassen 1997; Stieff and Wilensky 2003; Prince 2004; Jacobson and Wilensky 2006).

In addition to the influence of the individual teacher, Becher and Trowler (2001) highlighted broader differences in the nature of intellectual enquiry across the disciplines that may also influence the teaching methods used. Information-transmission orientations have been shown to be particularly common in science teaching (Handal, Lauvås, and Lycke 1990), which may reflect the accretionary, 'information-heavy', nature of the knowledge domain or be a consequence of a greater need to explain abstract concepts that will lie outside learners' prior experience. Whilst science teaching involves substantial active learning in laboratory classes, the traditional 'passive' lecture remains common and the place where students will likely develop understanding of the subject's conceptual structure.

The influence of presentation software such as PowerPoint in potentially guiding knowledge representations has been recognised. Beyond accusations that it can amplify an information-transmission style of teaching for some tutors (Jones 2003; Tufte 2003; Ward 2003), more significant are concerns about the influence that the software templates can exert on how a teacher develops representations of the discipline knowledge structure (Adams 2006; Kinchin 2006; Tufte 2006). The presentation template defaults to a format of subheadings and bullet points which can draw lecturers down a strongly linear, sequential teaching pathway that supports a cognitive style inconsistent with higher-thinking skills (Tufte 2003).

In the traditional lecture format, the process of selecting and sequencing material means that the lecturer can impart a linearity to knowledge structure that may inhibit student learning, which can become amplified by use of PowerPoint template designs (Kinchin, Chadha, and Kokotailo 2008). The lecturer selects many dimensions of the subject and discards many other branches and linkages, in effect choosing an

optimal route through the content. Kinchin, Chadha, and Kokotailo (2008) argue that PowerPoint invites the lecturer to reduce content further to bullet-point lists.

In summary, where the aim is to develop students' understanding of complex knowledge domains, the models of teaching applied will be significant in influencing how students encounter and develop their representations of the subject's conceptual structure.

Much teaching in natural sciences involves developing students' understandings of complex and dynamic, highly linked and interdependent systems. Such complex interactions cannot be represented easily as packages of information presented serially in chunks, either as chapters sequentially in a textbook or as a linear sequence of slides and bullet points in lectures.

Concept-mapping: developing complex learning

A growing body of research over the past two decades has recognised the benefits of concept-mapping techniques to develop students' understandings of more complex knowledge schemas in ill-structured subject areas (Clayton 2006; Spector 2006; Novak and Cañas 2007; Hay, Kinchin, and Lygo-Baker 2008). Introduced by Joseph Novak in 1972 (Novak 1998), concept maps are graphical tools typically in the form of a web diagram that can be used in multiple ways in education to aid planning, learning, synthesis or for assessment.

It is recognised that a key design principle for learning complex systems is to make the organising conceptual framework explicit (National Research Council 2000; Jacobson and Wilensky 2006). This builds on the understanding that experts not only do things differently from novices, but also *see* things differently (Lesh 2006), so making visible the interconnections between conceptual elements in a knowledge framework can better enable learners to build a mental picture of the subject structure. This is a key application of concept-mapping, enabling relationships and pathways between knowledge structures to be organised and represented, usually with links between concepts represented by arrows connecting various concept and sub-concept 'nodes' (Novak and Cañas 2007).

Hay, Kinchin, and Lygo-Baker (2008) explained how concept-mapping methods could add significantly to the quality of teaching at university level when introduced alongside conventional teaching and learning strategies, encouraging students to consciously and deliberately relate newly learnt knowledge to prior and other knowledge. It is the process of allowing the student to experience the complexity phenomena through the mapping activity that is considered essential to triggering any conceptual change necessary to develop cognitive 'expertise' (Jacobson and Wilensky 2006). Hay, Kinchin, and Lygo-Baker (2008) also highlighted the value of concept-mapping in prompting reflection on knowledge construction processes and how complexity in understanding develops.

Evaluating the impact of concept-mapping

Some researchers have evaluated the impact of concept-mapping interventions qualitatively by examining students' perceptions of the value of the activity (Gold and Coaffee 1998; Marangos and Alley 2007). Quantifying enhancements to learning has proven much more difficult. Ruiz-Primo and Shavelson (1996) explain that much quantitative performance assessment in science focuses on yielding evidence about

what students can do in terms of following procedures in investigations and in providing accurate solutions to a problem. They suggest there is a big inferential leap in going from observations of performance to the cognitive processes and higher-order thinking used by students in carrying out the investigation and pointed to a lack of research evidencing whether such inferential leaps were supported empirically. In providing insights into students' knowledge structures, concept-mapping techniques they argue, can provide an effective means to develop and evaluate aspects of students' cognition, such as their ability to organise and inter-relate subject concepts.

In examining the use of concept maps for assessing learning, Ruiz-Primo and Shavelson (1996) did, however, recognise a number of problems with the method. They identified that a wide variety of techniques and scoring methods were reported in the literature that were not linked to any integrative theory. In addition, they also questioned the validity of claims that concept maps were providing reliable scoring mechanisms. In addition, Kinchin (2001) points out that relatively few of the early evaluations of concept-mapping applications compared quantitatively measured outcomes or were subject to peer-review by publication.

Evaluation models tend to reflect the purpose of the activity. In many cases, concept maps have been used as a mechanism to measure development in learning over time, typically comparing the structure and complexity of participants' concept maps drawn at the beginning and the end of the learning activity (Hay 2007) or progressively over the duration of a course (Pearsall, Skipper, and Mintzes 1997).

Other studies have adopted experimental designs and measured impact quantitatively by comparing student performance between control and experimental groups. In these studies, the dependent variable, performance, is usually measured as course-test scores (Fraser and Edwards 1985; Rooda 1994). Test scores are commonly utilised as a quantitative measure of the quality of student learning, but as Ruiz-Primo and Shavelson (1996) point out, caution is needed to prevent over-interpretation in making inferences about the cognitive processes (see also research into effectiveness of journal writing on learning quality by Connor-Greene 2000). A particular problem with using simplistic test scores alone as a measure of impact is that they usually measure a totality of knowledge so inferences about the cognitive level will be contaminated by other factors such as writing quality, breadth of knowledge or general completeness, for example.

Shepard (2000) argues that, beyond measuring individual student-learning outcomes, a fundamental role for any formal or formative assessment activity should be the opportunity to examine and improve teaching practices, so evaluations and assessments that reveal insights into the cognitive processes in addition to the outcomes are essential in developing assessments that ultimately enhance learning.

This evaluation adopts a quasi-experimental design to compare learning outcomes of student cohorts who were exposed to a concept-mapping intervention with a cohort that was not. In this study, the outcomes are quantitative measures of the structural complexity in students' thinking expressed as written learning outcomes. In addition, these measures are compared to simpler generic measures of overall performance.

Evaluation of concept-mapping in an earth sciences programme

Teaching context and rationale

This investigation evaluates the impact of a concept-mapping intervention within a Level 3, year-long module 'Geohazards and Applied Geoscience', part of the geology

and physical geography programme at Liverpool John Moores University in the UK, a large modern university with over 20,000 undergraduate and 4000 post-graduate students. The module examines the causes of geological hazards and the strategies employed to prevent or mitigate their effects, with teaching and assessment grading shared between two tutors. Throughout the course, the two-hour weekly teaching sessions comprised a mix of lectures interspersed with question and answer discussion, and practical problem-solving exercises where students completed predefined tasks.

A key topic on the course was the study of landslide hazards, as a good example of a complex natural system. Landslide instabilities are rarely caused by a single, simple causal factor but by the interactions of a range of quasi-static and dynamic variables that change through time. Students are required to develop understanding of how short- or long-term changes in one system (e.g. climate or land use), will impact on other systems (soil structure, groundwater, vegetation, erosion). These natural systems are typically inter-related and in a state of balance, with changes in one system variable causing changes to impact across to others. Developing students' understanding of the interconnected 'web' of causes is pivotal in determining their ability to design landslide hazard prediction models, and was a intended key learning outcome for this part of the course.

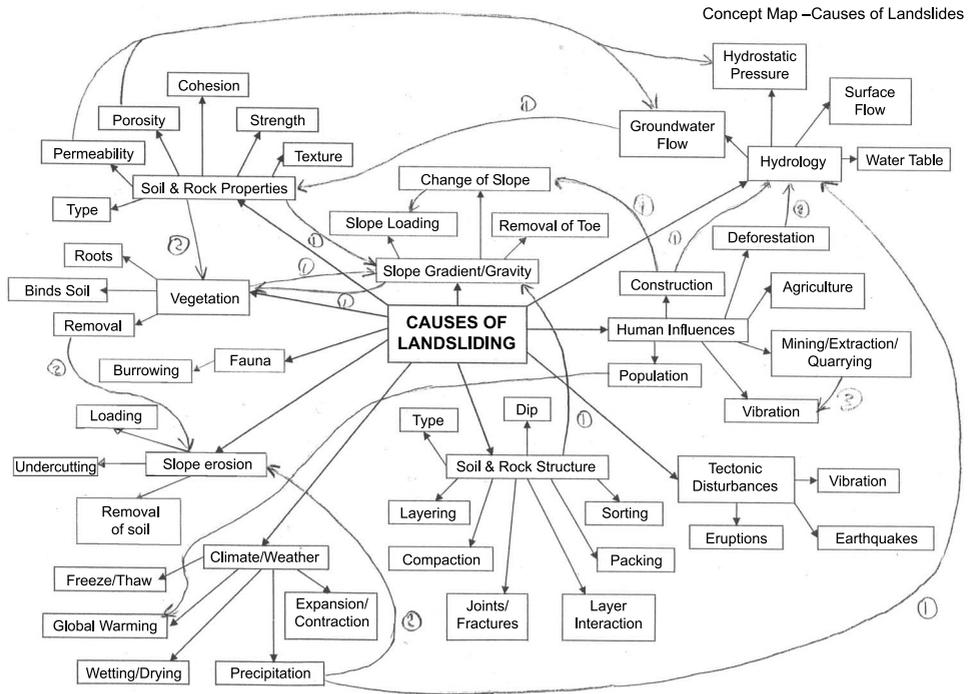
Concept-mapping intervention

Prior to the academic year 2006–2007, the topic 'landslides' had been taught through traditional, largely didactic methods. It had become apparent to the tutors from marking students' coursework that, despite good evidence of core subject knowledge, many students were struggling to integrate key concepts in this part of the course with a tendency for many to present knowledge in disconnected and unsophisticated ways rather than explore the critical interactions.

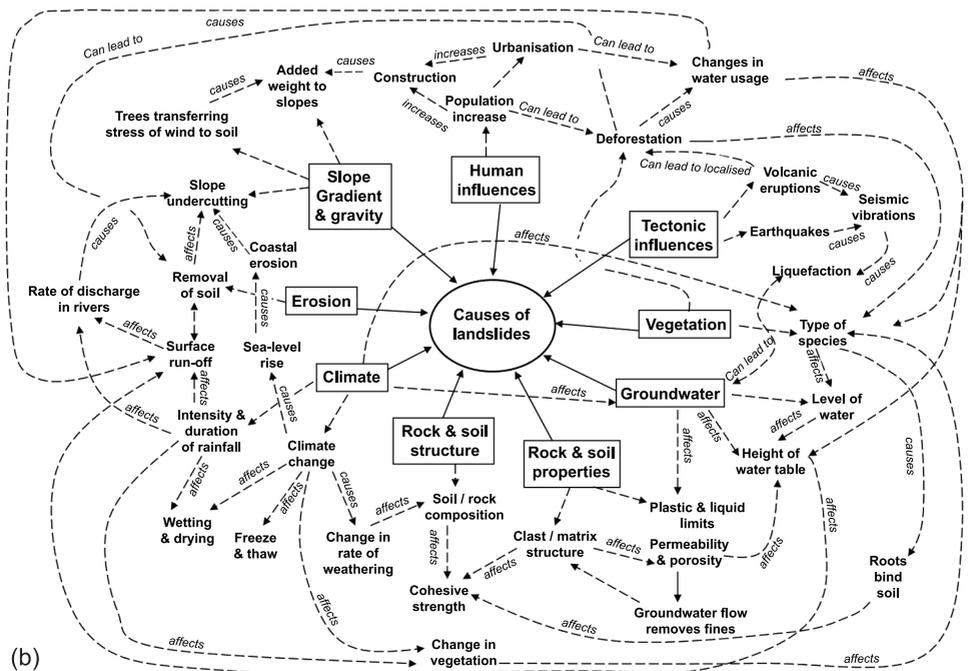
To address this shortcoming, it was decided in the academic year 2006–2007 to introduce a concept-mapping exercise. In the first of two sessions, students were given a large blank sheet of paper onto which they drew a central box with the title 'Causes of Landslides' and a series of 'nodes' linked by spokes to the central title box to represent different causal factors to be discussed. Through an accompanying lecture, a range of causal factors were briefly introduced and elaborated through class discussion exploring how each new factor might impact on or be affected by the other. During the discussion, students expanded the relevant concept node on the map into sub-branches and represented linkages. At the close of the session, they were asked to take away their map to complete in their own time and to bring this along the following week, where discussion continued. Figure 1a shows an example of a student map.

The majority of students took well to the activity although some struggled with the actual drawing of the map due its developing complexity, so a summary version was made available (Figure 1b). The main objective of the exercise was to examine the relationships between concepts, not simply to produce a concept map, but the diagram allowed interconnections to be represented explicitly which would have been difficult through note-taking alone. The intervention is defined not only as the physical construction of the concept map, but also includes the class discussion that formed a key part of the exercise.

After the lectures, a two-day field trip was run to observe and examine these landslide concepts first hand, this was followed by the production of a field report assignment (Assignment A, worth 15% of the module mark) in which students



(a)



(b)

Figure 1. (a) Example of student-drawn concept map. Circled numbers describe the nature of link as follows (1) to determine gradient, (2) to influence vegetation type, and (3) to trigger for failure; (b) summary concept map drawn by tutor and distributed at the second session.

reported their interpretations of the causes of landsliding at the study site. The remainder of the module studied different geological hazards and included two further assignments: a literature review (Assignment B, worth 15% of the module mark), and in Semester 2 a survey report (Assignment C, worth 20%) plus an end-module exam, worth 50%. Coursework items were formally graded using a dedicated marking rubric by the tutor who set the particular assignment task, and for each assessment a minimum sample of eight scripts were randomly selected and blind second-marked as part of the departmental procedures to ensure consistency in grading. In these procedures, if a difference greater than five percentage points arises between the grades awarded by the two markers, the work is discussed and a new mark agreed, although this situation did not arise for the sample groups studied.

Research aims and design

The aim of this evaluation was to measure the impact of the mapping intervention on student-learning outcomes in the assignment that followed the exercise. The study is a mixed method design comprising two strands. Strand 1 utilised an action research, quasi-experimental pre-test, post-test design to assess the significance of differences in learning outcomes between a 'control' cohort, who were not exposed to concept-mapping, and groups who participated in the exercise after its introduction to the module. This strand set out to test the following null hypotheses:

H1: There is no significant difference in the structural complexity of learning outcomes reflected in coursework between student groups who participated in the concept-mapping exercise and student groups who did not.

H2: There is no significant difference in the grade profile of coursework between student groups who participated in the concept-mapping exercise and student groups who did not

Strand 2 aimed to elicit students' views on the usefulness of the concept-mapping exercise to illuminate the quantitative study. It comprised a simple two-item survey administered at the end of the course.

Sample

The sample comprised three classes of undergraduate students taking the third-year module 'Geohazards and Applied Geosciences' in consecutive years: academic years 2005–2006 (Group 1), 2006–2007 (Group 2) and 2007–2008 (Group 3). The concept-mapping intervention was introduced in the academic year 2006–2007, so Group 1 serves as a 'control' group with Groups 2 and 3 as 'experimental' groups. With the exception of the intervention, teaching and learning were otherwise comparable in terms of the teaching personnel, content, delivery and assessment methods over the study period.

The sample cohorts included only those students who completed the module and all coursework assignments (Table 1). The three groups are of unequal size, which is not an ideal condition for a comparison, but this is an applied research study located in a 'live' setting, and such limitations are inevitable when dealing with natural variations in class size.

Table 1. Profile of the three sample cohort groups.

Module: Geohazards and Applied Geoscience	Number	Gender	
		M	F
Group 1 ('Control' group) Prior to introduction of concept-mapping exercise	28	20	8
Group 2 Participated in concept-mapping exercise	17	9	8
Group 3 Participated in concept-mapping exercise	24	16	8
Totals	69	45	24

Sample group comparison

The field report (Assessment A) examined the topics explored through the concept-mapping applied to a case-study site. Any impact of the intervention was likely to be manifest in the students' learning outcomes expressed in this assignment. Subsequent courseworks (Assignments B and C) were deemed unlikely to be influenced by the concept-mapping exercise due to the specific demands of these tasks. As such, the performance profiles for Assignments B and C can serve as a test for comparing the similarity of the three sample groups outside the intervention.

Performance data for Assignments B and C are shown in Table 2. Mean grades for the two assignments were similar. Performance differences were tested statistically. A one-way ANOVA was used to test for differences among the three cohorts for both assignments. Results showed that there was no significant difference among the mean grades of the three groups for Assignment B, $F_{(2,66)} = .245, p > .05$ nor for Assignment C, $F_{(2,66)} = 2.221, p > .05$. The assumption of equality of variance among the three groups was not violated, and thus the analysis results can be considered valid.

The differing mix of male to female students between the groups (Table 1) had no influence: *t*-tests reveal that gender differences in performance for both assignments were not statistically significant; $t_{(67)} = -1.82, p > .05$ for Assignment B, and $t_{(67)} = .343, p > .05$ for Assignment C.

Therefore, the comparison of the grades for Assignments B and C, deemed unlikely to be influenced by the concept-mapping exercise, suggests that performance profiles are comparable across the three groups. Any differences in the learning outcomes associated with Assignment A, which would likely be influenced by the

Table 2. Comparing performance profiles for the three groups for tasks unaffected by the concept-mapping intervention.

Cohort	Assignment B Literature review		Assignment C Survey report	
	Mean grade (%)	SD	Mean grade (%)	SD
Group 1 <i>n</i> = 28	55.86	9.96	54.18	10.90
Group 2 <i>n</i> = 17	57.88	7.36	54.88	8.86
Group 3 <i>n</i> = 24	56.67	9.99	58.83	8.93

exercise, should, therefore, be a reliable indicator for the impact associated with this intervention.

Study procedures and results

Strand 1: Did the concept-mapping exercise enhance students' learning?

Strand 1 of the study aimed to determine differences between the 'control' group and the 'experimental' groups relating to: (1) the structural complexity of thinking, expressed as written learning outcomes in Assignment A, and (2) the overall performance for this task expressed as grade.

Testing hypothesis 1

The Structure of the Observed Learning Outcomes (SOLO) taxonomy developed by Biggs and Collis (1982) is considered an appropriate model for measuring complexity of students' thinking expressed through their written learning outcomes. This hierarchical model allows classification of responses into levels of cognitive and structural complexity. The taxonomy has been applied in many studies as a framework to measure the higher-order conceptual structure of students' thinking expressed in essays or as verbal responses (Watson et al. 1995; Cambell, Smith, and Brooker 1998; Burnett 1999; Chan, Tsui, and Chan 2002; Munowenyu 2007). The taxonomy had been used in the assessment of this module for a number of years as an aid to marking and informing student feedback.

The model comprises five levels of cognitive complexity, prestructural, unistructural, multistructural, relational and extended abstract, with each category representing an increase in the number of organising and relational dimensions, as explained in Table 3. Finer sub-categorisation of the key multistructural and relational categories was added to eradicate conceptual ambiguity identified in relation to the taxonomy, following the study by Chan, Tsui, and Chan (2002). The addition of sub-levels allows for more refined characterisation of students' responses in these key areas (Table 3).

The rationale for introducing the exercise was to enhance students' understanding of the connectivity between course concepts. In terms of the SOLO taxonomy model, connectivity in thinking is represented by learning that falls into the relational or extended abstract categories.

Alongside the formal marking process, students' responses were coded by the author to one of the SOLO sub-categories shown in Table 3. A coding rubric was used that identified the number of concept variables represented and the extent to which the connections were drawn between them, identified via signifier words or phrases such as 'causes', 'influences', 'affects' and 'impacts upon'. Each report was assigned an appropriate coding score. During final analysis, to mitigate potential bias in the coding process, scripts from the three cohorts were collated and interdispersed prior to coding by a second independent rater using the same coding rubric. The scores assigned to each particular script by the two raters were then compared to determine the level of agreement.

Table 3 (right-hand half) shows the number of instances where a script was assigned to each SOLO category by the two raters for the three sample cohorts. Further detail is included in the Appendix 1 which shows the instances where Rater 2 agreed or disagreed with the score assigned to a particular script by Rater 1. Almost all variations

Table 3. SOLO taxonomy with sub-levels (modified after Chan, Tsui, and Chan 2002) used to characterise the level of structural complexity in students' work (Assignment A).

SOLO taxonomy level	Descriptions of student response	SOLO coding score	Group 1 <i>n</i> = 28		Group 2 <i>n</i> = 17		Group 3 <i>n</i> = 24	
			R1, R2	Mean	R1, R2	Mean	R1, R2	Mean
Prestructural	Question not understood; no relevant information	0						
Unistructural	Only mentions one relevant piece of information or variable	1	2, 1	1				
Multistructural/low	Picks up 2 or 3 independent aspects that are related to the topic but without further elaboration	2	4, 5	3	1, 1	1		
				2				
Multistructural/medium	Picks up a number of related pieces of information but presented serially or in isolation with no connections between concepts or ideas	3	3, 2	1			0, 1	
Multistructural/high	Picks up many related aspects and elaborates each, but with no connections between concepts	4	1, 2	1				1
				1	0, 1			3, 2
Relational/low	Connections drawn between variables and concepts in one or two parts of the assignment	5	8, 6	5	6, 8	4	7, 7	5
				3		5		
Relational/medium	Connections drawn between variables and concepts in many parts of the assignment	6	3, 7	4	6, 3	2	8, 6	4

Table 3. (Continued).

SOLO taxonomy level	Descriptions of student response	SOLO coding score	Group 1 <i>n</i> = 28		Group 2 <i>n</i> = 17		Group 3 <i>n</i> = 24	
			R1, <i>R2</i>	Mean	R1, <i>R2</i>	Mean	R1, <i>R2</i>	Mean
Relational/high	Overall generalisation of major concepts showing high levels of integration throughout assignment	7	6, 4	4	4, 3	3	3, 6	3
Extended abstract	Consistent generalisation and synthesis of concepts throughout and high level critical analysis.	8	1, 1	1	0, 1	1	3, 2	2
	Mean SOLO Score (Rater 1/Rater 2)		4.68, 4.75	4.71	5.65, 5.47	5.56	5.83, 5.83	5.83
	Percentage at relational level or higher (Rater 1/Rater 2)		64.3, 64.3	64.3	94.1, 88.2	91.2	87.5, 87.5	87.5

Notes: The columns on the right-hand side of the table show the numbers of student reports assigned to each SOLO category for each group by Raters 1 and 2 (*italics*) and the mean number of scripts allocated to each SOLO score/half-score when averaged between the two raters. The total mean SOLO score for each group is shown at the foot of the table in bold typeface.

in rating occurred between the sub-levels within the main SOLO categories. An explanation for this is that key signifiers for coding, such as the frequency and range of topics represented or frequency of indicator words for relational-level thinking (e.g. 'causes', 'affects'), make it relatively easy to assign work to the main category levels. Within the main SOLO categories, however, at the level of the sub-category divisions, agreement between the raters was less consistent (see Appendix 1), particularly in the relational category. An inter-rater reliability analysis using the Kappa statistic was performed to determine consistency among raters. As this is ordinal data, a linear-weighted kappa was used (k_w) as this generates a more realistic measure, ensuring that areas of minor disagreement are penalised less than areas where discrepancies are greater (Sim and Wright 2005; Viera and Garrett 2005). The inter-rater reliability measure was found to be $k_w = 0.797$ ($p < .001$), 95% CI (0.717, 0.877). This value represents 82% of the maximum attainable linear-weighted kappa for the table concerned (Appendix 1), so represents good reliability between the two raters.

For comparisons between groups and subsequent analysis, a mean rating score was used, the average of the number of instances scripts were allocated to each SOLO score/half score between the two raters. Findings in Table 3 reveal that for Group 1, who were not exposed to concept-mapping, 64.3% of the work was judged to be at the relational level or higher. The mean SOLO score for Group 1 was 4.71, equivalent to the upper multi-structural level. In contrast, for Groups 2 and 3 who participated in concept-mapping exercise prior to the assignment, the great majority were achieving a relational level or higher (91.2% and 87.5% for Groups 2 and 3, respectively) with SOLO score means equivalent to the low to mid relational level.

Testing was carried out to determine whether these differences in structural complexity were statistically significant. However, the assumption of equality of variance among the three groups was found to have been violated meaning that any testing using a one-way ANOVA would yield an inaccurate p -value with the probability of a false-positive being substantially higher than 5%. In such cases, particularly where sample sizes are unequal and on the small side, the most common alternative is the Welch ANOVA. This is a more robust and reliable measure that uses an F -statistic with degrees of freedom adjusted for unequal variances (Welch 1938; for review see Day and Quinn 1989). The test found significant variance between the three groups (Welch $F = 3.113$, $p < 0.05$). Planned comparisons tests confirmed that the significant difference lay between the control and experimental groups (Contrast 1, significant at $p = 0.025$) as indicated in Table 3. There was no statistically significant difference between the two experimental groups (Contrast 2, $p = 0.489$, not significant). The concept-mapping can therefore be shown to have made a statistically significant impact on the structural, cognitive complexity of students' learning.

Testing hypothesis 2

Comparison of the grade profiles for Assessment A was conducted to see if the impact extended to overall performance. Overall performance was judged using generic marking criteria used across the programme that considers a range of attributes applied to essays and scientific reporting. These include the quality of the introduction and rationale, the accuracy of data presentation and analysis, the level of interpretation, the coherency of writing and the general standard of presentation. Grading was expressed as a percentage point, with 40% equivalent to the pass level and 70% to first-class work. Performance data for the three groups are shown in Table 4.

Table 4. Performance profiles for Assignment A, the field report.

Cohort	Assignment A Field report Mean grade (%)	SD
Group 1 <i>n</i> = 28	56.14	9.26
Group 2 <i>n</i> = 17	60.82	9.54
Group 3 <i>n</i> = 24	60.58	8.47

The mean grade was four percentage points higher for the concept-map Groups 2 and 3, so the effect did appear to extend to performance. To determine whether this difference was significant statistically, a one-way ANOVA was conducted after confirming equality of variance. Testing found that the differences between mean grades across the three groups were not significant, $F_{(2,66)} = 2.090$, $p > .05$.

Therefore, whilst the concept-mapping exercise did have a significant effect on the level of structural complexity of students' work, the effect was not as strong in enhancing overall performance. This is not unexpected as performance was a function of a wider range of attributes, many of which would not have been affected by the intervention. Whilst students in the concept-mapping groups were demonstrating greater connectivity in relating course concepts, this was not necessarily paralleled by improvements in accuracy of interpretation or the general construction of the report.

Strand 2: Did students perceive the concept-mapping exercise to be useful to their learning?

At the end of the module, students were invited to participate in a short survey to reflect on the perceived usefulness of the concept-mapping exercises to their learning. Two questions were administered at the end of a revision activity in the final lecture of the module for Groups 2 and 3. Multiple choice questions were presented to students that they could respond to anonymously using an electronic interactive classroom voting system. Results are shown in Figures 2 and 3.

The majority found concept-mapping to be useful or very useful (Figure 2) with no one signalling a negative viewpoint. Responses to the second question, relating to how the exercise was deemed useful, were split between the two groups. Group 2 placed more value on the maps as an information-synthesising tool and Group 3 valued them more as a relating tool (Figure 3). This difference is difficult to explain but could reflect subtle differences in how the activity was introduced to the students. Although the teaching was comparable across the three years, in terms of the subject content covered, the teaching staff involved and the sequencing of lectures and practical tasks, there will be an inevitable degree of variation in how activities are explained and how discussion develops. The intention of the concept-mapping exercise had been to encourage recognition of relationships between concepts. The high proportion of responses recognising the value as a synthesising 'tool' could well be explained by the timing of this survey, during the revision period ahead of the exams where students were more likely to be referencing the exercise as a memorisation tool. The value of concept maps and diagrammatic representations serving to aid memorisation has been

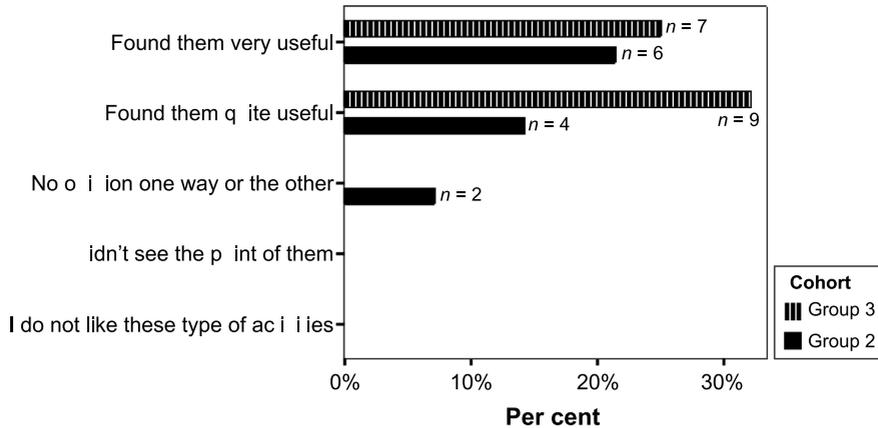


Figure 2. Responses to the question ‘Which of the following best describes your view towards the use of concept maps on this course?’ Group 2, $n = 12$; Group 3, $n = 16$.

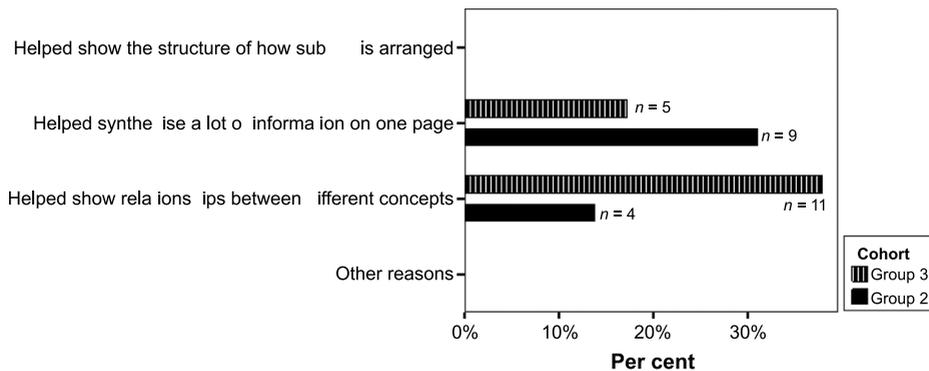


Figure 3. Responses to the question ‘If you found the concept maps useful, what was the main reason?’ Group 2, $n = 13$; Group 3, $n = 16$.

recognised previously from cognitive studies (Larkin and Simon 1987; Zhang and Norman 1994).

Interestingly, no students commented that the maps helped show how the structure of the topic was arranged (Figure 3). A key rationale for introducing the exercise was to develop the students’ understanding of the knowledge structure, to show the complexity in terms of connectedness of concepts and discourage artificial linear representations of knowledge. As the large majority of students who participated in concept-mapping were subsequently producing work at a relational level, this suggests the exercise was effective at developing an understanding of knowledge structure at an implicit level, but clearly from the students’ survey responses not at an explicit level. However, a key aim of the exercise was to demonstrate that the subject knowledge did not form a neat organised structure.

In addition, to this survey, there was some anecdotal evidence worthy of note. After the module exam, a number of students (Group 2) commented that they had used the technique as a revision aid to summarise information. On the examination scripts,

a few students had included simple concept maps as part of essay-plan notes prior to writing their essay. These observations suggest that some students had found a use for the maps beyond their initial intended use.

Discussion

It has been argued that there is a need to rethink our methods for teaching and assessing learning if we are to develop students' abilities, mental models and attitudes for learning complex subject matter (Jacobson and Wilensky 2006; Spector 2006; Knight and Page 2007).

At one level, this study has demonstrated the influence of a small-scale active learning method in improving the success to which students connect and integrate course concepts at a higher, relational level. This is significant because in doing so, students will be recognising, explicitly or implicitly, that there is a need to articulate connections and relations between concepts – to join-up thinking.

At another level, this can mark a bigger shift for students in how the broader subject itself can be understood. Amundsen, Weston, and McAlpine (2008) argue that concept-mapping is effective because a key way in which disciplines can be understood is through examination of their basic knowledge structure and relationships. Most systems in the natural world are not neat, simple hierarchies of knowledge, but complex, interconnected and interdependent networks where relatively small changes in one system can lead to wide-ranging impacts. So explicit and visual examination of these knowledge structures and connections can ultimately improve how students conceptualise their subject.

So if concept-mapping is so potentially valuable to science teaching, why are we not all doing it? Kinchin (2001) asked this question in relation to biology teaching at secondary school and concluded that part of the problem has been our obsession with curriculum development, not on how we as teachers teach and learn. From interviews and a review of the literature, Kinchin (2001) concluded that the biggest barrier to widespread adoption of effective active learning methods such as concept-mapping is the general lack of reflection by teachers on their own teaching practices and as a consequence, their epistemological beliefs about the teaching of their subject. He argued that simply trying concept-mapping as a teaching activity can be sufficient to prompt teachers' reflection on practice.

This is supported by Amundsen, Weston, and McAlpine (2008) who studied the use of concept maps by university academics as part of a series of course design workshops. They concluded that participants found the process of examining and making explicit course designs through mapping valuable both as a prompt to rethink course content and to focus on the relationships within the course structure. Both the studies by Kinchin (2001) and Amundsen, Weston, and McAlpine (2008) remarked that some teachers were initially sceptical about the concept-mapping technique, but ultimately came to value the activity.

It is also important to point out that, despite caution or scepticism on the part of teachers in adopting such methods, the students questioned in this study clearly found the activity to be of value to their learning and thus greater attention to the student voice should be heeded from such evaluations. This aligns with research by Marangos and Alley (2007) who reported that students clearly recognised and valued the learning benefits brought on by concept-mapping, particularly in identifying linkages and aiding understanding and memorisation of course concepts.

Finally, in their critique of research into higher education academics' beliefs about teaching, Kane, Sandretto, and Heath (2002) question the direct way in which belief systems are inferred to drive practice and suggest the role of the contexts of teaching and learning have been overlooked. This study has highlighted a key area where the *context* has been critical in driving the pedagogic approach, in the learning of complex subject material, in the need to develop students' higher-level thinking skills at mature stages of an undergraduate degree programme, and in the deployment of an appropriate measurement tool to characterise this learning.

Conclusion

This evaluation has provided further empirical evidence to support the positive impact of concept-mapping, in terms of developing higher-order, relational-level learning outcomes and recognising the value students place on such interventions in developing their learning. This is particularly significant in the context of university teaching in the natural, physical and social sciences in light of advances in our understandings of system complexity.

In particular, this study has demonstrated the need to measure enhancements to student learning using an appropriate assessment tool, in this case a modified version of the SOLO taxonomy of Biggs and Collis (1982). It has demonstrated how reliance on simpler performance measures such as test or coursework grades can underestimate the impact of the technique.

Acknowledgements

I wish to thank Dr Elizabeth Worth for assistance in coding data and for comments in preparation of this manuscript. I am also grateful for the very constructive comments from anonymous referees on the earlier draft.

Notes on contributor

Martyn Stewart is coordinator for pedagogic research at Liverpool John Moores University, based in the institution's Academic Enhancement Unit since 2000. Previously he held research and lecturing posts in structural geology at the Universities of Durham and Southampton. In addition to his current educational development work, he maintains an active teaching role in geology. His current research interests are in student engagement and active learning.

References

- Adams, C. 2006. PowerPoint, habits of mind, and classroom culture. *Journal of Curriculum Studies* 38: 389–411.
- Amundsen, C., C. Weston, and L. McAlpine. 2008. Concept-mapping to support university academics' analysis of course content. *Studies in Higher Education* 33: 633–52.
- Barnett, R. 2000. *Realizing the university in an age of supercomplexity*. Buckingham: The Society for Research into Higher Education/Open University Press.
- Becher, T., and P. Trowler. 2001. *Academic tribes & territories*. Buckingham: Society for Research into Higher Education/Open University Press.
- Biggs, J., and K.F. Collis. 1982. *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Bowden, J., and F. Marton. 1998. *The university of learning: Beyond quality & competence*. London: RoutledgeFalmer.
- Burnett, P.C. 1999. Assessing the structure of learning outcomes from counselling using the SOLO taxonomy: An exploratory study. *British Journal of Guidance & Counselling* 27: 567–81.

- Cambell, J., D. Smith, and R. Brooker. 1998. From conception to performance: How undergraduate students conceptualise and construct essays. *Higher Education* 34: 449–69.
- Chan, C.C., M.S. Tsui, and M.Y.C. Chan. 2002. Applying the Structure of the Observed Learning Outcomes (SOLO) taxonomy on students' learning outcomes: An empirical study. *Assessment & Evaluation in Higher Education* 27: 511–27.
- Chi, M.T.H. 1992. Conceptual change within and across ontological categories: Implications for learning and discovery in science. In *Minnesota studies in the philosophy of science: Cognitive models of science*, Vol. XV, ed. R. Giere, 129–86. Minneapolis: University of Minnesota Press.
- Chi, M.T.H. 2005. Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences* 14: 161–99.
- Clayton, L.H. 2006. Concept-mapping: An effective, active teaching-learning method. *Nursing Education Perspectives* 27: 197–203.
- Connor-Greene, P.A. 2000. Making connections: Evaluating the effectiveness of journal writing in enhancing student learning. *Teaching of Psychology* 27: 44–6.
- Day, R.W., and G.P. Quinn. 1989. Comparison of treatments after an analysis of variance in ecology. *Ecological Monographs* 59: 433–63.
- Entwistle, N., and P. Walker. 2000. Strategic alertness and expanded awareness within sophisticated conceptions of teaching. *Instructional Science* 28: 335–61.
- Fraser, K., and J. Edwards. 1985. The effects of training in concept-mapping on student achievement in traditional classroom tests. *Research in Science Education* 15: 158–65.
- Gold, J.R., and J. Coaffee. 1998. Tales of the city: Understanding urban complexity through the medium of concept-mapping. *Journal of Geography in Higher Education* 22: 285–96.
- Goldstone, R.L. 2006. The complex systems see-change in education. *Journal of the Learning Sciences* 15: 35–43.
- Handal, G., P. Lauvås, and K. Lycke. 1990. The concept of rationality in academic science teaching. *European Journal of Education* 25: 319–32.
- Hay, D. 2007. Using concept maps to measure deep, surface and non-learning outcomes. *Studies in Higher Education* 32: 39–57.
- Hay, D., I.M. Kinchin, and S. Lygo-Baker. 2008. Making learning visible: The role of concept-mapping in higher education. *Studies in Higher Education* 33: 295–311.
- Hmelo-Silver, C.E., and R. Azevedo. 2006. Understanding complex systems: Some core challenges. *Journal of the Learning Sciences* 15: 53–61.
- Jacobson, M.J., and U. Wilensky. 2006. Complex systems in education: Scientific and educational importance and implications for the learning sciences. *Journal of the Learning Sciences* 15: 11–34.
- Jenkins, A. 2004. *A guide to the research evidence on teaching–research relations*. York: The Higher Education Academy
- Jenkins, A., M. Healey, and R. Zetter. 2007. *Linking teaching and research in disciplines and departments*. York: The Higher Education Academy.
- Jonassen, D.H. 1997. Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Education Technology Research & Design* 45: 65–94.
- Jones, A.M. 2003. The use and abuse of PowerPoint in teaching and learning in the life sciences: A personal overview. *Bioscience Education E-Journal* 2–3. <http://www.bioscience.heacademy.ac.uk/journal/vol2/beej-2-3.aspx> (accessed January 10, 2010).
- Kane, R., S. Sandretto, and C. Heath. 2002. Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research* 72: 177–228.
- Kember, D. 1997. A reconceptualisation of the research into university academics' conceptions of teaching. *Learning & Instruction* 7: 255–75.
- Kinchin, I.M. 2001. If concept-mapping is so helpful to learning biology, why aren't we all doing it? *International Journal of Science Education* 23: 1257–69.
- Kinchin, I.M. 2006. Concept-mapping, PowerPoint, and a pedagogy of access. *Journal of Biological Education* 40: 79–83.
- Kinchin, I.M., D. Chadha, and P. Kokotailo. 2008. Using PowerPoint as a lens to focus on linearity in teaching. *Journal of Further & Higher Education* 32: 333–46.

- Knight, P., and A. Page. 2007. The assessment of 'wicked' competencies. Report to the Practice-based Professional Learning Centre, Open University, <http://kn.open.ac.uk/public/getfile.cfm?documentfileid=10623> (accessed January 10, 2010)
- Larkin, J.H., and H.A. Simon. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11: 65–100.
- Lesh, R. 2006. Modelling students modelling abilities: The teaching and learning of complex systems. *Journal of the Learning Sciences* 15: 45–52.
- Marangos, J., and S. Alley. 2007. Effectiveness of concept maps in economics: Evidence from Australia and the USA. *Learning & Individual Differences* 17: 193–9.
- Munowenyu, E. 2007. Assessing the quality of essays using the SOLO taxonomy: Effects of field and classroom-based experiences by 'A' Level geography students. *International Research in Geographical & Environmental Education* 16: 21–43.
- National Research Council. 2000. *How people learn: Brain, mind, experience and school* (Expanded ed.). Committee on Developments in the Science of Learning and Committee on Learning Research and Educational Practice, ed. J.D. Bransford et al., Washington, DC: National Academy Press.
- Novak, J.D. 1990. Concept-mapping: A useful tool for science education. *Journal of Research in Science Teaching* 27: 937–49.
- Novak, J.D. 1998. *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations*. Mahwah, NJ: Lawrence Erlbaum.
- Novak J.D., and A.J. Cañas. 2007. Theoretical origins of concept maps, how to use them and uses in education. *Reflecting Education* 3: 29–42.
- Pearsall, N.R., J.E.J. Skipper, and J.J. Mintzes. 1997. Knowledge restructuring in the life sciences: A longitudinal study of conceptual change in biology. *Science Education* 81: 193–215.
- Prince, M. 2004. Does active learning work? A review of the research. *Journal of Engineering Education* 93: 223–31.
- Prosser, M., K. Trigwell, and P. Taylor. 1994. A phenomenographic study of academics' conceptions of science learning and teaching. *Learning & Instruction* 4: 217–31.
- Rooda, L. 1994. Effects of mind mapping on student achievement in a nursing research course. *Nurse Educator* 19: 25–7.
- Ruiz-Primo, M.A., and R.J. Shavelson. 1996. Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching* 33: 569–600.
- Shepard, L.A. 2000. The role of assessment in a learning culture. *Educational Researcher* 29: 4–14.
- Sim, J., and C.C. Wright. 2005. The kappa statistic in reliability studies: Use, interpretation and sample size requirements. *Physical Therapy* 85: 257–68.
- Spector, J.M. 2006. A methodology for assessing learning in complex and ill-structured task domains. *Innovations in Education and Teaching International* 43: 109–20.
- Stieff, M., and U. Wilensky. 2003. Connected chemistry: Incorporating interactive simulations into the chemistry classroom. *Journal of Science Education and Technology* 12: 285–302.
- Trigwell, K., M. Prosser, and P. Taylor. 1994. Qualitative differences in approaches to teaching first-year university science. *Higher Education* 27: 75–84.
- Tufte, E.R. 2003. *The cognitive style of PowerPoint*. Cheshire, CT: Graphics Press.
- Tufte, E.R. 2006. *The cognitive style of PowerPoint: Pitching out corrupts within*. 2nd ed. Cheshire, CT: Graphics Press LLC.
- Viera, A.J., and J.M. Garrett. 2005. Understanding interobserver agreement: The kappa statistic. *Family Medicine* 37: 360–3.
- Ward, T. 2003. I watched in dumb horror. *Education Guardian*, May 20: <http://education.guardian.co.uk/egweekly/story/0,5500,959058,00.html> (accessed January 10, 2010).
- Watson, J.M., K.F. Collis, R.A. Callingham, and J.B. Moritz. 1995. A model for assessing high order thinking in statistics. *Educational Research and Evaluation* 1: 247–75.
- Welch, B.L. 1938. The significance of the difference between two means when the population variances are unequal. *Biometrika* 29: 350–62.
- Zhang, J., and D.A. Norman. 1994. Representations in distributed cognitive tasks. *Cognitive Science* 18: 87–122.

Appendix 1

Matrix showing agreement between raters of SOLO codes for student scripts. The table shows the instances where Rater 2 agreed with the score assigned to a particular script by Rater 1.

		Rater 1								Total	
		SOLO score									
		1	2	3	4	5	6	7	8		
Rater 2	SOLO score	1	1								1
		2	1	4	1						6
		3		1	1	1					3
		4			1	3	1				5
		5					14	6	1		21
		6					6	9	1		16
		7						2	10	1	13
		8							1	3	4
Total		2	5	3	4	21	17	13	4	69/69	

Notes: Rater 1 = main researcher; Rater 2 = independent verifier.

Copyright of Assessment & Evaluation in Higher Education is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.