Assessment Tools for Teaching and Learning Technical Report #43

COGNITIVE PROCESSES IN asTTle: The SOLO TAXONOMY

Abstract: The Structure of Observed Learning Outcomes (SOLO) cognitive processing taxonomy, developed in the 1970s and 80s by two Australian academics—John Biggs and Kevin Collis, categorises mental activity by quantity and quality attributes of the activities required by students or by the observable products of student work. This taxonomy has been used in asTTle to categorise student performance on every task in Reading/Pänui and Mathematics/Pängarau. This report explains the SOLO taxonomy and its psychological basis, and provides examples of using SOLO in assessment and education in general. Other technical reports will examine the performance of students by the SOLO taxonomy categories.

Submitted by the Assessment Tools for Teaching and Learning team,

Auckland UniServices Ltd

University of Auckland

August 2004

COGNITIVE PROCESSES IN asTTle: The SOLO TAXONOMY

This report provides an explanation of the SOLO taxonomy used to ascertain cognitive processing in asTTle assessment questions and tasks. asTTle is funded by the Ministry of Education to Auckland Uniservices Ltd. at the University of Auckland to research and develop an assessment application for Reading, Writing, Mathematics, Pänui, Pängarau, and Tuhituhi for Years 4-10 (Levels 2-6) for New Zealand schools. We acknowledge this funding, and thank the Ministry of Education for their continued assistance in the development of this project.

The Structure of Observed Learning Outcomes (SOLO) cognitive processing taxonomy, developed in the 1970s and 80s by two Australian academics—John Biggs and Kevin Collis, categorises mental activity by quantity and quality attributes of the activities required by students or by the observable products of student work. This taxonomy has been used in asTTle to categorise student performance on every task in Reading/Pänui and Mathematics/Pängarau. This report explains the SOLO taxonomy and its psychological basis, and provides examples of using SOLO in assessment and education in general. Other technical reports will examine the performance of students by the SOLO taxonomy categories.

We have based parts of this report on our experience of teaching the SOLO taxonomy at many asTTle item writing, reviewing, and signature workshops and to various education classes and groups at the University of Auckland. This work, including the examples, was refined with the constructive feedback of Mr Earl Irving, Ms Sheryll McIntosh, and Associate Professor Roger Peddie (lecturers on the University of Auckland Education 224/225 paper).

John Hattie

Project Director, asTTle

August 2003

The bibliographic citation for this report is:

Hattie, J.A.C., & Brown, G.T.L. (2004, September). *Cognitive processes in asTTle: The SOLO taxonomy*. asTTle Technical Report #43, University of Auckland/Ministry of Education.

Cognitive Processes in asTTle: SOLO Taxonomy

Table of Contents

Cognitive Processes in asTTle: SOLO Taxonomy

In a recent book on education in the knowledge age, Carl Bereiter (2002) used Popper's three worlds to make sense of much of what we strive for in school: the physical world, the subjective or mental world, and the world of ideas. There are major parallels with the three worlds of achievement: surface knowledge of the physical world, the thinking strategies and deeper understanding of the subjective world, and the ways in which students construct knowledge and reality for themselves as a consequence of this surface and deep thinking. This third world, often forgotten in the passion for teaching facts and thinking skills, is entirely created by humans, is fallible but capable of being improved, and can take on a life of its own. Students often come to lessons with already constructed realities (third worlds), which if we as teachers do not understand and assess before we start to teach, can become the stumbling block for future learning. If we are successful then the students' constructed realities (based on surface and deep knowing) are the major legacy of teaching. It is certainly the case, as Bereiter documents, that "much of what is meant by the shift from an industrial to a knowledge society is that increasing amounts of work are being done on conceptual objects rather than on the physical objects to which they are related" (p. 65).

The major shift, therefore, needs to be from an over reliance on surface information (world 1), a misplaced assumption that the goal of education is deep understanding (world 2; e.g., the development of thinking skills), towards a balance of surface and deep learning leading to student's more successfully constructing defensible theories of knowing and reality. The major purpose of this report is to present a model for ensuring that achievement assessments in schools have at least a balance of items at both the surface and deep levels, such that the students are primed to the construction of knowing and exploring realities for themselves.

It is noted that most secondary students in NZ schools take a surface approach to understanding both how and what they should learn (Brown, 2002a); whereas their teachers claim that the goal of their teaching is enhancing deep learning (Brown, 2002b). Brown (2002a) found that the majority of Year 11 students defined studying or learning with surface strategies or methods (i.e., revision, re-reading, and reviewing of the year's work) and strongly agreed that learning involved building up knowledge by getting facts and information. In contrast, teachers preferred a deep view of learning, usually focused on academic, cognitive development, while at the same time, emphasising surface methods of teaching in order to prepare students for high-stakes qualification examinations or assessments (Brown, 2002b). This emphasis on surface approaches means that students tend to experience very few opportunities or demands for deep thinking in contemporary classrooms.

We also commence by commenting on the nature and quality of teachers' questioning of student learning, which has been examined in some detail. Teachers daily ask many questions of students, with some studies reporting hundreds of questions per day by teachers (Marzano, 1991; Wood, 1991). A large percentage of teacher questions have been identified as requiring simple recall of factual knowledge; for example, Gall (1970) claimed that 60% of teachers' questions required factual recall, 20% were procedural, and only 20% required thought by the students. Other studies have found the proportion of surface thinking questions is in the order of 80% plus (Airasian, 1991; Barnette, Orletsky, & Sattes, 1994; Gall, 1984; Kloss, 1988). Teachers' questioning may not elicit deep thinking from students because they understand questioning is how teachers lead and control classroom activity; in other words, students know that the teacher already knows the answer to the questions and so do not think about the answers at all (Gipps, 1994; Torrance & Pryor, 1998; Wade & Moje, 2000). So much of daily classroom life is "knowledge telling". There is good evidence that when teachers implement rich, divergent, higher order thinking questioning based on an understanding of the surface knowledge as part of their classroom repertoires enhanced learning takes place (Black, Harrison, Lee, Marshall, & Wiliam, 2003; Gall, 1984; Wood, 1988). This enhanced learning relates to the more defensible constructions of reality (world 3).

The key issue addressed in this report is how to devise an assessment model that values a balance of surface and deep processing. The key for the asTTle development is the use of a defensible taxonomy of processing -- the SOLO taxonomy. Like most taxonomies, SOLO describes the processes involved in asking and answering a question on a scale of increasing difficulty or complexity.

It is acknowledged that the most well known taxonomy in education is Bloom's taxonomy. That taxonomy refers to the type of thinking or processing required in completing tasks or answering questions; that is, know, comprehend, apply, analyse, synthesise, and evaluate (Bloom, Engelhart, Furst, Hill & Krathwohl, 1956). Appendix A outlines many of the difficulties of the Bloom Taxonomy, and for now it is only important to note that the original developers of Bloom's Taxonomy have recently abandoned their model primarily because of major deficiencies in this simplistic and incorrect hierarchy of six steps; and our claim is that these deficiencies are best resolved using the SOLO model

When using the SOLO taxonomy, either the questions would be written in a different manner, or the test scorer would concentrate on classifying the responses only. An example of re-writing to maximise the correspondence between the question asked and the answer expected is seen in Figure 1.

Figure 1
SOLO Taxonomy Applied to Questions/Tasks about Picasso's *Guernica*

Unistructural. Who painted *Guernica*?

Multistructural. Outline at least two compositional principles that Picasso used in *Guernica*.

Relational. Relate the theme of *Guernica* to a current event.

Extended Abstract. What do you consider Picasso was saying via his painting of *Guernica*?

The SOLO Taxonomy

The approach to cognitive processes adopted in the Assessment Tools for Teaching and Learning (asTTle) has been based on the Structure of Observed Learning Outcomes (SOLO) Taxonomy (Biggs & Collis, 1982; Collis & Biggs, 1986). The SOLO Taxonomy was developed by analysing the structure of student responses to assessment tasks in response to given body of information or knowledge and identifying the type of thinking exhibited by extended written responses. SOLO has been applied in many different school subjects: poetry (Biggs & Collis, 1982), history (Biggs & Collis, 1982), mathematics (Collis & Romberg, 1992), geography (Courtney, 1986; Stimpson, 1989), science (Collis & Davey, 1986), economics (Pong, 1989), chemistry (Holbrook, 1989), computer studies (Ki, 1989), and assessing attitudes towards teenage pregnancy (Kryzanowski, 1988).

The taxonomy consists of two major categories each containing two increasingly complex stages: surface and deep (Surface = Unistructural and Multistructural; Deep=Relational and Extended Abstract) The taxonomy makes it possible, in the course of learning, teaching, or assessing a subject to identify in broad terms the level at which a student is currently operating. In the simplest language the SOLO taxonomy consists of four levels: one idea, multiple ideas, relating the ideas, and extending the ideas: one, many, relates, and extend.

Figure 2

SOLO Taxonomy Category Definitions

Unistructural. One aspect of a task is picked up or understood serially, and there is no relationship of facts or ideas.
Multistructural. Two or more aspects of a task are picked up or understood serially, but are not interrelated.
Relational. Several aspects are integrated so that the whole has a coherent structure and meaning.
Extended Abstract. That coherent whole is generalised to a higher level of abstraction.

The two surface level responses involve understanding of ideas or facts. Unistructural responses and questions require the knowledge or use of only one piece of given information, fact, or idea, obtained directly from the problem. With an increase of quantity, multistructural responses or items require knowledge or use of more than one piece of given information, facts, or ideas, each used separately, or in two or more distinct steps, with no integration of the ideas. In contrast, the two deep processes constitute a change of quality of thinking that is cognitively more challenging than surface questions. Relational responses or questions require integration of at least two separate pieces of given knowledge, information, facts, or ideas, which when working together answer the question. In other words, relational questions require learners to impose an organising pattern on the given material. The highest level of the SOLO taxonomy, extended abstract requires the respondent to go beyond the given information, knowledge, information, or ideas and deduce a more general rule or proof that applies to all cases. In this latter case, the learner is forced to think beyond the given and bring in related, prior knowledge, ideas, or information in order to create an answer, prediction, or hypothesis that extends the given to a wider range of situations.

The Psychological Basis of the Four Levels.

Biggs and Collis (1982) based their model on the notion that in any “learning episode, both qualitative and quantitative learning outcomes are determined by a complex interaction between teaching procedures and student characteristics” (p. 15). They emphasised the roles played by: the prior knowledge the student has of the content relating to the episode, the student's motives and intentions about the learning, and the student's learning strategies. As a consequence, the levels are ordered in terms of various characteristics: from the concrete to the abstract, an increasing number of organising dimensions, increasing consistency, and the increasing use of organising or relating principles. It was developed to assess the qualitative outcomes of learning in a range of school and college situations and in most subject areas; hence the title of the taxonomy: Structure of the Observed Learning Outcome.

There are four major ways that the four levels can increase in complexity:

<u>Capacity</u> Each level of the SOLO taxonomy increases the demand on the amount of working memory or attention span. At the surface (unistructural and multistructural) levels, a student need only encode the given information and may use a recall strategy to provide an answer. At the deep (relational or extended abstract) levels, a student needs to think not only about more things at once, but also how those objects inter-relate.

<u>Relationship</u> Each level of SOLO refers to a way in which the question and the response interrelate. A unistructural response involves thinking only in terms of one aspect and thus there is no relationship possible. The multistructural level involves a many aspects but there is no attention to relationship between these aspects. At the relational level, the student needs to analyse and

identify an appropriate relationship between the many ideas, and at the extended abstract level, the student needs to generalise to situations not experienced or beyond the given environment.

Consistency and closure. These refer to two opposing needs felt by the learner. On the one hand, the student wants to come to a conclusion and thus answer or close the question. But on the other hand, the student wants to experience consistency so that there is no contradiction between the question posed, the material given, and the answer provided. Often, when there is a greater need for closure, less information is utilised resulting in an answer or response is that is less consistent. In contrast, when a high level of need for consistency is required, a student may utilise more information when conceiving an answer, but may not be able to reach closure if external factors do not permit. At the unistructural level, the student often seizes on immediate recall information, but at the extended abstract level, the student must integrate potentially inconsistent ideas and must tolerate the possibility of inconsistency across contexts.

Structure The unistructural response takes one relevant piece of information to link the question to the answer. The multistructural response takes several pieces and links them to the question. The relational response identifies and makes use of an underlying conceptual structure and the extended abstract requires a generalised structure such that the student demonstrates an extension beyond the original given context.

One of the major attractions of the SOLO model is that is suggests a basis for progressing students up the four levels. Thus, there are clear implications for how teachers can develop programs that enable students to enhance the depth of their learning. Students should be assisted to advance in the following ways:

From ”knowing nothing” to unistructural.

The teacher needs to help students 'join the game' with its new rules and its different way of conceptualising reality. For example, when teaching to read, it is worthwhile to take advantage of the children’s imaginations; that is, to use their interest in listening to stories and extracting meaning from pictures.

From unistructural to multistructural.

The teacher needs to concentrate on consolidating and automating the unistructural knowledge and skills, building a store of knowledge, and encouraging students 'to do more' with their knowledge base.

From multistructural to relational.

This involves more than “getting to know more about a topic or being adept at following through a sequence of procedures; it includes understanding or integrating what is known into a coherent system wherein the parts are inter-related. This interrelationship comes about as a result of an ability to form an over viewing principle which can be derived from the information given” (p. 196).

From relational to extended abstract.

This process requires dedicated hard work to master abstract concepts and relationships which allows the student to derive more generalised principles and transfer understanding to new tasks and situations.

An Illustration of SOLO in Action

A large-scale application of the SOLO taxonomy is demonstrated via the asTTle (University of Auckland & Ministry of Education, 2003) assessment software that is now in use in New Zealand classrooms. The cognitive processing most likely involved in answering each reading, pānui, pāngarau, and mathematics assessment item has been classified within asTTle. Meagher-Lundberg and Brown (2001) and Thomas, Holton, Tagg and Brown (2003) describe in reading and mathematics respectively, how the SOLO category for each item was agreed on by at least two-thirds majority rule by panels of teachers. In asTTle V4 about 45% of the reading and mathematics items were classified as deep while 30% of the pānui and pāngarau items were identified as deep (Table 1).).

Table 1

asTTle V4 items by SOLO by Subject

Subject	Surface	Deep	Total
English Medium			
Reading	828 (52%)	772 (48%)	1600
Mathematics	844 (58%)	622 (42%)	1466
Maori Medium			
Pānui	333 (62%)	208 (38%)	541
Pāngarau	503 (78%)	145 (22%)	648

When the application creates any test, it is a requirement that each created test has at least 25% surface and 25% deep items in each 40-minute paper-and-pencil test. asTTle then reports student performance by surface and deep items compared to appropriate year and sub-group norms. The next sections of this Report outline how the items were written to ensure that each measured one level of the SOLO taxonomy.

Writing items with the four levels of SOLO

1. *An example from Reading.*

Throughout the asTTle development teachers have been involved in workshops writing and reviewing assessment tasks according to the SOLO taxonomy. In reading and panui, where the curriculum emphasis is on close or comprehension reading of texts, the SOLO taxonomy was found to be powerful in distinguishing differences in items (Meagher-Lundberg & Brown, 2001). A typical workshop session commences with an example of how to write items at each of the four levels from a short story. For example, consider the students have just had the classic children's fairy tale of *Goldilocks & the Three Bears* read to them, and that the teacher wishes to explore and extend children's understanding of the author's use of imagery. A simple unistructural question (one idea) that could be posed is "*Whose house did Goldilocks go into*?" Answering this question requires simple attention to the word house and matching to a single idea provided early in the story, that the Bears owned the house. Such a task clearly involves recall or knowledge, to use Bloom's term, of the single fact, knowledge, or idea: Bears ⇔ house.

An example of multistructural (many ideas) is to ask "*What are three aspects about the way the bears live that tell us that the story is not a real life situation?*" It would be anticipated that listeners would identify factors relating to the anthropomorphisation of animals (i.e., bears don't live in houses, sleep in beds, eat cooked food, bears don't talk, etc.). The response requires simple listing of human traits that can be seen in the way the story portrays the life of bears.

A further advance in cognitive complexity and processing can be found by requiring respondents to identify the relationship underlying the data given "*Goldilocks eats the baby bear's food, breaks his chair, and sleeps in his bed. What does this tell us about the kind of person she is*"*?* In this case, the list of evidence or details is fore grounded by the teacher and requires the child to create some sort of meaningful relationship based on the given information. Because relational questions may be somewhat more challenging, the teacher may find that providing a list of alternatives is useful to guide or scaffold the thinking of the learner. In this question, acceptable answers could include *Goldilocks is bad, naughty, or destructive* and the teacher may then ask the students what that evidence tells us about the kind of person the baby bear is, which would also require relational thinking.

The extended abstract approach is to seek a pattern across a wider range of cases including those not given specifically by the Goldilocks story itself. Such a question is "*Why do nursery tales allow wild animals to act in human fashion?*" Again to ease students' approach to such a challenging question for which there may be many correct answers, the teacher may find it useful to provide a list of possible answers, such as:

a) to reveal negative things about human character in a safe way;
b) to show the oneness of nature and humanity;
c) to entertain children who easily believe that wild creatures can talk;
d) to give children courage to face dangers of adult life.

Although this list may NOT have one clear correct answer, the list does provide the opportunity for extensions beyond the basic ideas and require extension to contexts much wider than the one given in the story. Such debate elicits extension and abstraction in that defending a position requires finding commonalities across both given and non-given, but related cases, and finding higher order principles that account for the differences in details and which also take into account much grander purposes for the use of literature than simple entertainment.

This Goldilocks example is built around a body of given knowledge (a famous story) and an interest in an educationally relevant learning objective (finding information, and making inferences in reading). The same approach can be utilised in other languages and subjects.

2. An example from Pangarau

asTTle has used the SOLO taxonomy in the development of assessment items for learning objectives in Maori-medium curricula (i.e., pānui, tuhituhi, and pāngarau). These items were developed independently of the English-medium items but with the same basic constraints; that is, they had to fit to curriculum objectives and levels and measure both deep and surface cognitive processing.

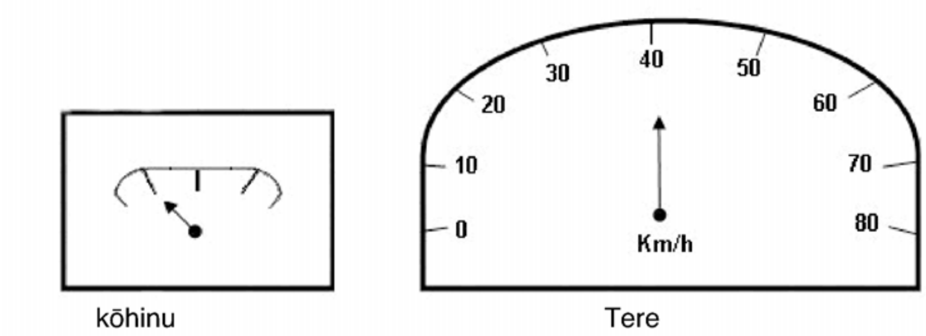
In the example testlet shown (Figure 3), students are asked to use the fuel gauge (kōhinu) and the speedometer (tere) to answer three questions. The first two questions assess the fraction of petrol remaining (1/4) and the current speed of the automobile (40 km/h) respectively. Both require surface cognitive processing in that the student has to serially locate the correct gauge, read the scale, and extract the correct value being pointed at.

The third question, on the other hand, asks the student to determine whether the car is fast or not and provide an explanation to support the position. Not only does this task require communicating mathematical ideas, but it also requires the student to think relationally. The student must use the maximum speed shown on the speedometer (80 km/h) and infer, based on knowledge of external, ungiven facts, that the car is relatively slow because the manufacturer did not supply it with a maximum value higher than 80 km/h. and motor vehicles generally cannot go as fast as the maximum speed shown on the speedometer.

Figure 3

asTTle Pangarau Testlet Showing Both Levels of SOLO Taxonomy

Koia nei te āwhata o tētahi motuka.



He aha te hautau o te kōhinu e toe ana?

1. ______________________________

He aha te tere o te motuka i tēnei wā?

2. ______________________________

Te āhua nei he motuka tere tēnei? Whakamārama mai i tō whakautu.

____

____

____

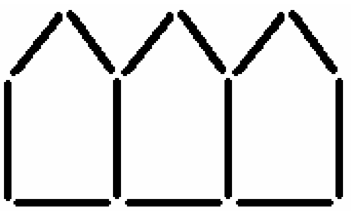
3. ______________________________

3. An example from Mathematics.

When teaching understanding patterns in number/algebra a common task is to provide students with a diagram of a pattern (e.g., house outlines made with match sticks—Figure 4). It is then possible to devise a series of questions that explore both the surface and deep thinking around the objects and principles involved in pattern making.

Figure 4

Matchstick Houses: Patterns in Number



Houses	1	2	3
Sticks	5	9	__

A simple <u>unistructural</u> question (one idea) requires elicitation of a response based on handling one aspect of the given data; *"How many sticks are needed for 3 houses?"* This task can be answered most simply by counting the number of sticks shown in the diagram to come up with the answer of 13. The next level, <u>multistructural</u>, is to require two or more ideas that are handled independently or serially. For example, *"How many sticks are needed for each of these three houses?"* requires the learner to take the given pattern and count the sticks for each house (5 each). To require deep thinking, the teacher needs to frame a question about finding a <u>relationship</u> within the given material, rather than persist with surface approaches of count or draw-and-count: For example, *"If 52 houses require 209 sticks, how many sticks do you need to be able to make 53 houses?"* (Answer: 213). In order to respond, a child must detect that for every additional house four more sticks are required, regardless of how many houses there are. <u>Extended abstraction</u> within the domain of algebra is a commonly achieved through explicit attention to more general rules that apply in all cases, whether such rules are expressed in words or algebraic terminology. Such an extended abstract task would be *"Make up a rule to count how many sticks are needed for any number of houses".* This demands a response that identifies not only the four sticks per house but also the need for one more to close off the last house in the series (e.g., $S = 4H + 1$). If a student provided this response, it would demonstrate understanding not only the relationship of sticks to houses but also the abstract extension that applies to all cases regardless of actual numbers.

4. An example from Writing.

The assessment of writing in asTTle requires teachers to evaluate the characteristics of student scripts against the progress indicators developed to show progress against Curriculum Levels 2 to 6 inclusive (Glasswell, Parr, & Aikman, 2001; Coogan, Parr, &

Hoben, 2003). Seven key dimensions of written work were identified as crucial in the development of writing for socio-culturally defined purposes of persuade, instruct, narrate, describe, explain, recount, and analyse.

The four deep dimensions are audience and purpose awareness, content selection, organisation, and language resources. These dimensions require the writer to integrate material provided in the writing prompt with the informational needs of an imputed audience, with his or her prior knowledge, and with his or her assigned goals or purposes in writing. This relational thinking is expected to result in an integrated response which selects appropriate content, sequences it according to purpose and audience, and expresses it with appropriate selection of language. Clearly this aspect of writing requires deep thinking.

The three surface dimensions (i.e., grammar, punctuation, and spelling) require from the writer the ability to handle language objects and rules in a multi-structural, sequential fashion. Simply put, the writer, when editing or proofing the written work, applies various known and given rules one at a time to ensure accuracy of the obvious features of written language. For example, when faced with the mis-spelled "cheif" a writer may apply the known rule 'i before e, except after c' without any need to integrate that rule with any other language content or rule; thus, elegantly exercising surface thinking as described by SOLO. It is not coincidental that these dimensions are commonly referred to as 'surface' characteristics of writing, which we are very happy to let computers check for us.

5. Constructing testlets.

A powerful assessment system that can be used by teachers, and is a well used aspect within asTTle, is to construct a series of 4 items relating to a single text or prompt, with each item measuring one level of the SOLO taxonomy. This series of items is called a testlet. Consider the following testlet which has four items, one at each level of the taxonomy (Figure 5).

Figure 5

A testlet constructed according to SOLO.

`O´	O	O	O	O	O	O	O	O	O
Sun	Mercury	Venus	Earth	Mars	Jupiter	Saturn	Uranus	Neptune	Pluto

Temperature: HOT--------------------COOL COLD

1. Which is the planet furthermost from the sun? (Unistructural)

2. Which two planets are closest to Earth? (Multistructural)

3. Explain how distance from the sun and temperature are related. (Relational)

4. Given the Earth's position relative to the sun, in what ways does this affect the Earth's climates and seasons? (Extended abstract)

In the first question, only a single piece of information is required. In the second question, the student is required to use two separate pieces of information to work out the answer. In the third question, it is necessary that the student sees the relationship between the given information about distance from the sun and planetary temperature. Finally, in the fourth question, the student has to go beyond the information provided in the item to deduce a more general principle as to the effects of the Earth's position to the sun and the effects on the climates and seasons.

6. Scoring open-ended questions

It is relatively easy to use the SOLO levels to identify and categorise student responses to open ended items. Figure 6 shows four different responses to a science question that a teacher has asked in a test or in a class session about reasons for darkness at night. The unistructural response shows a simple cause and effect explanation focused on a simple understanding of the phenomenon. The multistructural response replicates the same simplistic quality of cognition but increases the number of causes identified. In contrast, the relational response identifies and elucidates the relationship of light and night with a second phenomenon of rotation. In the extended abstract response, the answer extends the relationship to the shape and axis of the planet as part of a generalised explanation for the phenomenon.

Figure 6

SOLO Classification of Responses to Open-Ended Items

Question: Why does it get dark at night?

Unistructural. Because the sun goes to the other side of the world.

Multistructural. Because the earth is spinning and the sun is going round the earth.

Relational. It gets dark at night because the sun goes around the earth once for 12 hours and for the other 12 hours it is day as the sun is around the opposite side of the earth.

Extended abstract. The earth is spherical in shape and rotates about its north-south axis. As it rotates, at any one time the half of the Earth's sphere facing the sun will be in light while the opposite half will be in shadow. As the earth is rotating continuously, a point on the earth's surface will pass alternately through the lighted half and the shaded half.

7. General principles for asking questions

From these examples, it can be seen that there are a number of general principles for asking questions – either by a teacher orally in a class during the development of a lesson, or by a teacher in a written assessment task. First, decide on a learning intention related to the given material. Second, find one element in the material, related to the learning intention, that students can identify, locate, or complete using one process or one piece of knowledge or data. Third, devise a question requesting a list of two or more things that the student can locate from the learning material, related to the learning intention. Fourth, provide a list of ideas and ask what they have in common to generate a response that requires finding relations between the ideas. Finally, decide on the general principle used in the relational question (i.e., what class of event, personality, situation, rule, etc. does this relationship in the given context connect to?) and devise a question, task, or activity that assists the student to identify such a general principle and thus require extended abstract thinking.

We suggest that advancing through these four levels is a defensible method for teaching material, mimics how many students learn more complex material, and obviously, can serve as a model for developing assessment items. It is important that there is a close match between the level of SOLO a question is designed to require and the cognitive processing actually used by the student to answer the question. Such an approach leads to excellent measurement, and also assist student to understand the intention of the question, and the level of detail and complexity required to answer the question.

It is also important to have confidence that teachers can classify questions into the four levels of SOLO in a consistent manner. In one study, Hattie and Purdie (1994) asked 30 teachers to classify 19 multiple choice items taken directly from Bloom et al. (1956). Half the group classified the items into Bloom levels and then into SOLO levels; the other half did the same but in counter-balanced order. The average accuracy was 60% correctly allocated into the exact SOLO levels and 96% at the level or one level different, and 40% into the exact Bloom levels and 75% at the level of one level different. Early in the development of asTTle, a panel of primary school literacy teachers were asked to classify the cognitive demand of about 300 reading comprehension assessment items according to both the Bloom and SOLO taxonomies (Meagher-Lundberg & Brown, 2001). The raters reached a high level of agreement using the SOLO taxonomy – 73% agreement. Biggs and Collis (1982) used two judges to code a series of history questions and reported a 79% agreement in categorising the responses into the correct level, 11% at one level difference, and only 9% at more than one level difference.

Another major advantage of the SOLO model is that it separates the concept of “difficulty” from “complexity”. While it may be usual for questions to increase in difficult as they increase in complexity, this is not always the case. For example, asking students to answer mathematics questions with bigger numbers or harder numbers (e.g., irrationals) or with more steps is not necessarily increasing the cognitive depth of student thinking (from surface to deep), although it may be more difficult. Depth is not the same as difficulty — perhaps it is this confusion that explains why so many questions posed by teachers do not require students to use higher order thinking skills but instead require a greater attention to details (McMillan, 2001).

Reporting surface and deep learning in asTTle

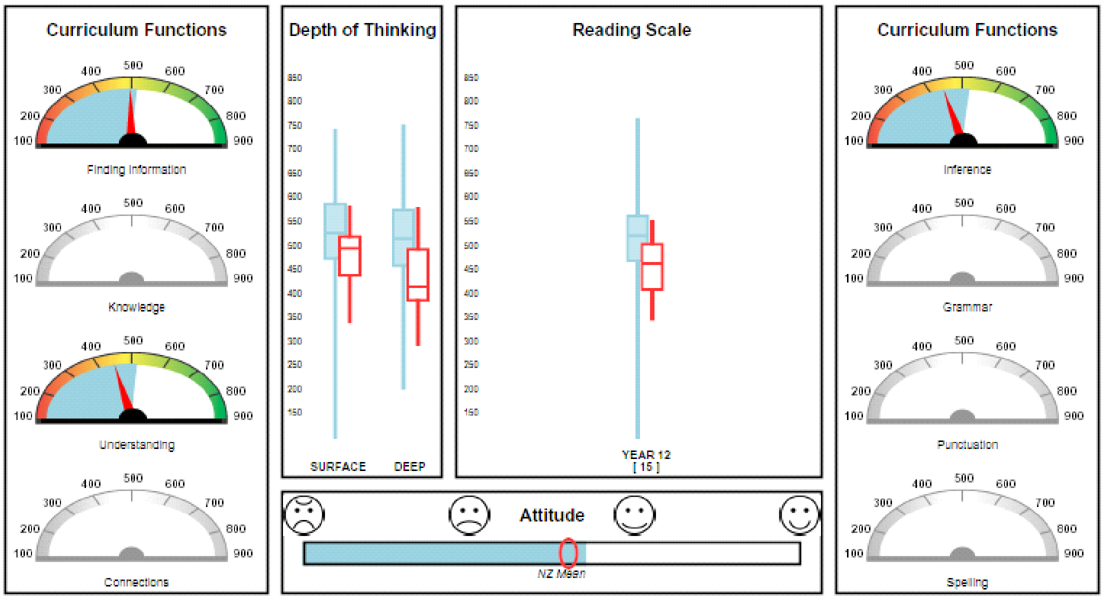
The asTTle scoring engine calculates each student's performance using item response theory techniques to determine their weighted aggregate performance across the items classified by teachers as requiring either surface or deep cognitive processes. The asTTle Console Report displays the cohort performance of students compared to that of national norms (see Figure 7). Every item in asTTle has been classified by a team of teachers at one of the four levels of SOLO. However, in the asTTle reports the first two levels (unistructural and multistructural) have been combined as surface thinking, and the latter two levels (relation and extended) have been combined as deep thinking. It is thus important for teachers and school leaders to make appropriate interpretations about the level of complexity mastered by their students. This section provides some guidance for using the reporting of cognitive thinking in asTTle.

1. Determine the relative strengths or weaknesses

For example, Figure 7 shows clearly that the median performance of students being reported in red is significantly lower for deep processing than their performance on surface processing, although the range of scores for deep and surface processing are almost identical. In such a case, a teacher may want to consider how to encourage the students to relate and extend the various knowledge of ideas they clearly possess. Encouraging the students to gain more knowledge (ideas) may not be as beneficial as learning how to relate and extend the ideas.

Figure 7

asTTle Console Report Showing Performance by SOLO Processing



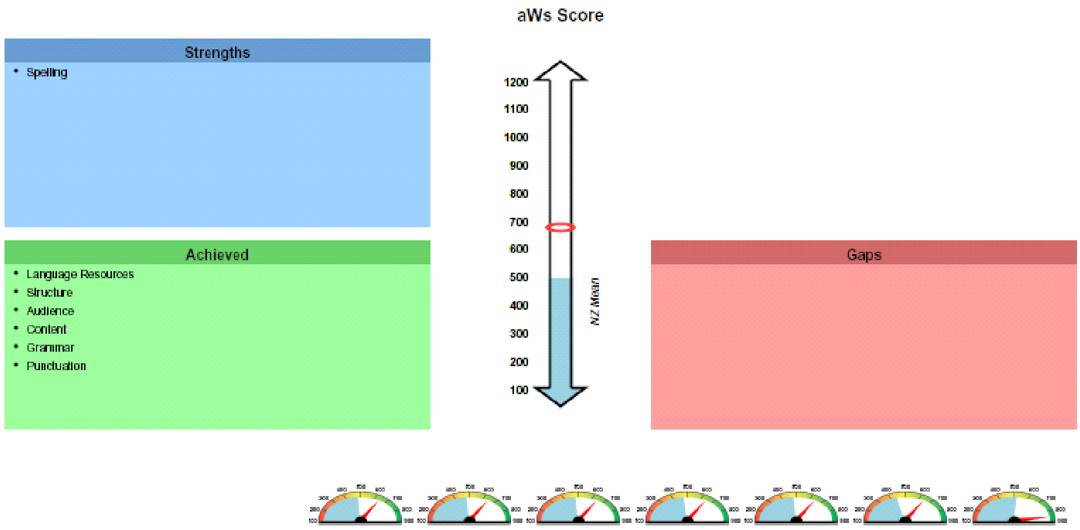
From this type of reporting it is expected that teachers will determine the relative strengths or weaknesses of students in handling the deep or surface processes within a subject, leading to decisions about appropriate instructional activities and resource selection. For example, in the course of teaching a group of Year 12 students doing an alternative English program, the teacher used asTTle to identify and focus her teaching program on students' ability to think deeply by drawing relational inferences from their close reading of text (Andersen, 2004). Subsequent monitoring found that, through the teacher's use of relational learning activities (e.g., reciprocal reading, asking meta-cognitive reflection questions, and using Three-Level reading guides to stimulate deep thinking) with a range of books, the deep thinking skills of the class on average reached the same level as the surface thinking skills some four months later (Andersen & Brown, in preparation).

2. *Mismatch between Processes: Surface Greater than Deep*

It is quite probable that students will exhibit greater competence at handling surface processing tasks than those that require deep processing. In the example Individual Learning Pathways Report for a writing task (Figure 8), it is clear that the student has achieved a very high degree of accuracy in spelling relative to both the curriculum levels (>4 Advanced) and the average for Year 6 students (858 compared to 524). Overall the student has scored within Level 4 for all other aspects of writing so the spelling score represents a personal strength within an a generally strong writing ability for the student. Note that the surface processing score has been raised by the high spelling score to the same level as the deep features, but the student's overall score has been raised only a little. In this case, the teacher has the option of extending the student's spelling ability, but greater educational achievement would come from focusing on the deep features which still lag behind the spelling ability. Thus, the interaction of both the cognitive processing score and the domain scores is useful to inform meaningful educational planning; after all, increasing surface processing ability at the expense of deep processing is a good way to train a proof reader or editor, not a writer.

Figure 8

Individual Learning Pathways Report Writing



	aWs	Surface	Deep	Audience	Content	Structure	Language Resources	Grammar	Punctuation	Spelling
This student	673	683	667	639	639	639	639	639	639	858
Level	4P	4P	4P	4B	4B	4B	4B	4B	4B	>4A
Year 6 mean	487	486	484	483	484	485	482	484	445	524

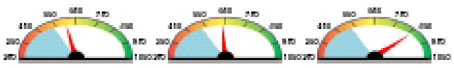
3. Mismatch between Processes: Deep Greater than Surface

It is possible for students to score higher on deep processes than surface processes. In the case shown in Figure 9, the student has scored significantly higher on the statistics items than either the measurement or number operations items and has gained a Level 5

Advanced for deep thinking compared to Level 4 Proficient for surface thinking. This indicates that the items measuring the measurement and number operations items were not only easier in terms of mathematical content but also easier in terms of the type of cognitive processing required. The statistics items obviously required explicitly identifying relationships or rules within data, while the operations and measurement items required a series of serially applied operations or mathematical calculations leading to a correct response. Thus, the student has the ability to process deeply but either is careless or inattentive when handling surface tasks (perhaps because they lack challenge) or else lacks key knowledge needed to accurately operate the surface requirements of the items. If the latter situation, it is vital that building up the missing knowledge or understanding or, perhaps fluency and automaticity, would be of great value to this student's educational progress.

Figure 9

Extract from Individual Learning Pathways Report Mathematics



	aMs	Surface	Deep	Number Operations	Measurement	Statistics
This student	814	710	905	605	638	855
Level	4A	4P	5A	3A	4B	5B
Year 7 mean	542	546	540	541	531	533

Other Uses of SOLO in Assessment and Evaluation

In addition to being used to shape and analyse educational assessments, SOLO has been used in a variety of other educational contexts and purposes. Five examples are provided to illustrate more fully the four SOLO levels and the value of the underlying model.

1. SOLO and Study skills

The first example is found in a meta-analysis of study skills programs by Hattie, Biggs & Purdie (1996). A major issue in that study was the power of the SOLO method to classify interventions. A unistructural study skills intervention is based on one relevant feature or dimension, such as an intervention focused on a single point of change, like coaching on one algorithm such as training in underlining, using a mnemonic device, or helping students to reduce anxiety. A multistructural intervention involves a range of independent strategies or procedures, but without any integration or orchestration concerning individual differences, or content or contextual demands. Examples are typical study skills packages which include learning a variety of strategies (such as note taking, organisation skills). A relational intervention occurs when all the components were integrated to suit the individual's self-assessment, were orchestrated to the demands of the particular task and context, or involved a degree of self-regulation in learning (e.g., meta-cognitive interventions emphasising self-monitoring and self-regulation, and many attribution retraining studies). An extended abstract intervention occurs when the integration achieved in the previous category was generalised to a new subject, or used in future learning.

Unistructural and multistructural programs were highly effective with virtually all students when studying material required only low level cognitive involvement (e.g., memorisation of specific information). Relational programs, integrating the informed use of strategies to suit the content were highly effective in all domains (i.e., academic performance, study skills, and affect) over all ages and ability levels, but were particularly useful with high ability students and older students.

2. Identifying expert teachers

The SOLO taxonomy has been used very powerfully to exhibit the impact expert teachers have on student learning in an American study on Nationally Board Certified Teachers (Bond, Smith, Baker, & Hattie, 2000). Students' persuasive writing essays were examined from the classes of teachers who had either passed or failed the National Board Certification examinations. Startlingly, 75% of student writing in non-certified teachers' classes exhibited surface cognitive processing, while 75% of student writing in Board certified teachers' classes revealed deep cognitive processing. This clear difference pointed to the vital difference between Board certified and non-certified teachers; expert teachers enable and require students to work at a deep cognitive level

The conclusion was that expert teachers are more likely to lead students to deep rather than surface learning. These teachers tend to structure lessons to allow the opportunity for deep processing, set tasks that encourage the development of deep processing, and provide feedback and challenge for students to attain deep processing.

3. Evaluating Gifted Programs

A third example is the evaluation of gifted programs by Maguire (1988). He used the SOLO taxonomy as part of an evaluation of programs for bright and gifted students in elementary and junior high school. The students in the program often pursued the objectives of the program by working independently on projects, working together in small groups, or participating in a mentorship program. This "diversity in learning activity may lead to uneven levels of knowledge about a particular content domain, in spite of the fact that levels of attainment of higher order objectives such as critical thinking may be uniformly high. In many situations the content provides a vehicle for instruction and may differ across students. It is not easy to find instruments that are relevant to program objectives, flexible enough to capture the creativity and divergence expected in performance from these kinds of students, yet at the same time possesses utility and validity" (p. 10). Thus, to evaluate the program, Maguire devised two writing and three mathematics tasks, and the answers to these questions were coded into the SOLO levels. It was expected that there would be more students in the gifted program at the higher levels compared with students of similar ability not in the program.

Maguire argued that the SOLO approach seemed to tap a complex of deep understanding, motivation, and intuition as applied to a particular task, thus it was appropriate to assess complex achievements, deep understanding, higher order skills, and strategic flexibility (cf., Snow, 1989). He found that students operating at the higher levels of the SOLO taxonomy (i.e., relational and extended abstract) tended to have higher scores on deep and achieving styles. Students who gave higher level responses to the SOLO writing tasks were also students who were more deeply engaged in their learning, while students who produced lower level products seemed to have more superficial approaches. When he compared the SOLO profiles from the students in the gifted program with a group of students in the regular classrooms identified as being gifted, and another group identified by the teachers as 'potentially gifted', there were no discernible differences. As Maguire concluded, the results provide "a picture of a program that is not yet succeeding" (p. 9). The use of the SOLO levels, however, allowed this researcher to "put outcomes on a common base while at the same time avoiding the confinement of standardized instruments. ... (SOLO) has been a very useful tool for detecting problem areas” (p. 9).

4. Evaluating teacher education students

At Richard W. Riley College of Education in South Carolina, the field placement of student teachers is evaluated using a SOLO taxonomy. Each observation is rated by the observer (an experience teacher, teachers’ college staff) according to the SOLO levels. Performance at both the Multistructural and Relational levels lead to satisfactory completion of the internship, as at these levels it is claimed that the teacher candidate demonstrates an ability to consider alternatives when making instructional decisions. Typically, these alternatives involve some of the following issues in balancing theory and practice: recognizing individual vs. group needs or interests, making concepts understandable or interesting to students, managing learning activities within time and space constraints, determining effective learning strategies for specific purposes, selecting activities that challenge students to do their best work. The five levels for the observational rubric are:

Prestructural = There may be preliminary preparation, but the task itself is not attacked in an appropriate way. Little evidence of organized thinking about classroom tasks. Planning and teaching are not well organized. Little reflection on students or experiences. The Prestructural level is not acceptable for teacher candidates working with students in classrooms. It generally shows lack of understanding or lack of ability to organize activities. In terms of professional behaviour, a teacher candidate at this level displays very poor and unacceptable behaviors in all of the following areas: communication, appearance, attitude and ethics. Please indicate this level by checking a 1 on the evaluation.

Unistructural = One aspect of the task is performed or understood serially. However, there is no relationship to other facts or ideas. Single-focused thinking about classroom tasks. Teacher implements lesson plans and activities in a rote fashion; cannot explain reasons for instructional choices or decisions. Reflections after teaching are focused mainly on events; i.e., what happened. This level represents a beginning level of functioning. The teacher candidate follows directions or implements activities in a way that shows procedural knowledge, but not understanding of ways to choose effective instructional activities. In terms of professional behaviour, a teacher candidate at this level displays very poor and unacceptable behaviors in one or more of the following areas: communication, appearance, attitude and ethics. Please indicate this level by checking a 2 on the evaluation.

Multistructural = Two or more aspects of the task are performed or understood serially with limited interrelationships to other ideas. The individual has a limited understanding of how concepts or ideas fit together. Considers more than one issue in making instructional choices. Lessons and activities are purposeful. Teacher can explain plans for lessons in terms of students or in terms of curriculum or theory in a limited way. Reflections after teaching focus on important events. The Multistructural level indicates that the teacher candidate considers one or two alternatives for making instructional choices (activities, lessons, or management) and can provide a reason for choices. In terms of professional behaviour, a teacher candidate at this level displays acceptable behaviors in the areas of: communication, appearance, attitude and ethics. Please indicate this level by checking a 3 on the evaluation.

Relational = Several aspects are integrated so that the whole has a coherent structure and meaning in and of itself. It is like fine woven fabric. Lessons and activities are purposeful and thoughtful. Sequences are clear. Teacher can explain lesson plans in terms of students and in terms of curriculum goals or theories. Reflections after teaching focus on meaningful events, significance of students' responses, or quality of learning. The Relational level indicates that the teacher candidate considers a variety of alternatives for instructional decisions, and weighs the potential effect of those choices. The candidate can explain choices in terms that show understanding of students and effective instructional strategies as they relate to curriculum goals or theories and his/her personal philosophy of teaching. In terms of professional behaviour, a teacher candidate at this level displays excellent behaviors in the areas of: communication, appearance, attitude and ethics. Please indicate this level by checking a 4 on the evaluation.

Advanced Expertise (Extended Abstract) = The coherent whole is raised to a higher level of performance showing advanced expertise. This level is not a standard expectation for undergraduate teacher candidates. This level is reserved for those teacher candidates who are operating more like experienced professional teachers than like teacher candidates. It is possible that a teacher candidate could behave like an experienced teacher in some categories. If you believe that a teacher candidate is operating at this level, please write in descriptors that would explain or provide evidence of an advanced performance.

5. In evaluating students learning in a classroom

Smith and Hattie (in preparation) are evaluating an instrument they devised to measure a teacher's depth of teaching, and a student's depth of understanding within any one lesson. The instrument is oriented around the four basic premises of the SOLO model: Capacity (How much working memory is required of students in this lesson? Or what thinking strategies (i.e., is recall sufficient?) are students required to use in this lesson?); Relationship (What is the relationship between the topic, assignment, lesson and the response?); Consistency (Do students need closure and consistency, or how well can student tolerate ambiguity? Or a continuum from immediate need for closure to a deep-seated concern to achieve beyond what is presented); and Structure (Does the student use relevant pieces of information or concepts to approach the task?). The teacher observation instrument is included in Appendix B.

Conclusions

SOLO is a hierarchical taxonomy based on an analysis of the structural characteristics of questions and answers. It identifies characteristics of increasing quantity and quality of thought, and it is critical to note that both surface and deep cognitive processes are needed when mastering school work; it is not the case that Surface is Bad, Deep is Good. It is a cliché, but it is difficult to be deep without some surface material to think deeply about. Students must be able to master both surface and deep thinking and they can gain such proficiencies if teachers require, through their questioning of learning and/or via the nature of the assessment tasks, students to develop both surface and deep thinking. The SOLO taxonomy and the item development strategy outlined in this article are powerful in ensuring that teachers' questions require students to think deeply about the ideas and material they are studying.

As noted above, the SOLO taxonomy not only suggests an item writing methodology, but the same taxonomy can be used to score the items, can allow for crediting partial knowledge and, most important, can be used for meaningful reporting to teachers and students. As demonstrated in the asTTle application, SOLO can also be used to address many vexing issues in measurement. For example, items in an 'item bank' can be classified according to the various levels of SOLO, and such information can assist in computer adaptive tests for choosing appropriate items to be administered, for ensuring that sufficient items at each level have been presented prior to stopping the adaptive testing, and for scoring such adaptive tests.

SOLO can be used for operationalising the quality of learning and standards for teachers and students to aim for with respect to particular tasks. For example, the teacher could prescribe a minimum of *u*% unistructural, *m*% multistructural, *r*% relational, and *e*% extended abstract for a particular course, class, or school. Each SOLO level is a metric of the complexity of the material, and thus it is easier for a teacher to select a task for the student relevant to the performance of the students, or even more desirable, one level higher than where students are currently performing. Teachers could be encouraged to use the 'plus one' principle when choosing appropriate learning material for students (Where to next?). That is, the teacher can aim to move the student one level higher in the taxonomy by appropriate choice of learning material and instructional sequencing.

The SOLO taxonomy offers teachers of students at all levels an alternative tool that can be used not only as a basis for selecting items for a test (as was the original intention of the Bloom taxonomy), but which also can provide a structure to help teachers devise appropriate instructional processes, engage in curriculum and task analysis, make judgements about the quality of learning that takes place in the classroom, and instigate appropriate remedial procedures where necessary.

The SOLO levels arise from an understanding of the process of student learning, and a concern to develop qualitative criteria of learning that have formative as well as summative value (Biggs & Collis, 1982). The power of the SOLO taxonomy to evaluate the quality as well as the quantity of learning is a feature absent from many of the procedures typically used by teachers to make judgements about student learning. In this respect, we have provided examples, broad in scope, of the capacity of the SOLO model to form the basis of measures of the quality of student learning.

A feature of the SOLO model that we have highlighted in this report is the way in which it reflects the complexity of human learning. Unlike the assumptions on which the Bloom taxonomy has been predicated, there is no separation between content and context, and there is recognition of the role of both the student and the teacher in student learning. The level at which the student is operating can be assessed, and the teacher can plan lessons and classroom activities aimed at helping students progress from simple unistructural responses to more complex ones involving relational and abstract thinking. We have provided examples from various content domains demonstrating the applicability of the SOLO taxonomy to a range of teaching aspects. In this report, we have not been able to outline the full range of possibilities of the use of SOLO in the various subject domains. Biggs and Collis (1982), however, provide many examples of how teachers can use the taxonomy not only to assess content knowledge, but to assist students to explore the full range of possibilities related to a given task. In poetry, for instance, teachers can use the taxonomy just as easily to explore students' knowledge of the structural features of a poem as they can to evaluate students' grasp of its metaphorical meaning, or to explore students' affective reactions. No less possible is the ability of a teacher of mathematics to use the taxonomy to devise instructional methods aimed at developing reasoning, creativity and positive attitudes in their subject.

In the conclusion to their book, Biggs and Collis (1982) noted that if their speculative suggestions were ultimately supported by research, then what started out as a descriptive model for a circumscribed context - school learning - might contain within it the seeds of a theory of learning with a wide range of application. Twenty years hence, there is ample evidence to demonstrate the power of the SOLO model in educational contexts in which age, subject matter, and instructional processes are widely varied. In this report we have presented tangible evidence of the capacity of the SOLO model to inform the practice of educators at all levels of student learning and, in particular, demonstrated and justified why the asTTle application is premised on assessing the learning growth from surface through to deep learning.

References

Airasian, P. W. (1991). *Classroom Assessment* (3rd ed.). New York: McGraw-Hill.

Andersen, I. (2004, July). *Flight paths through literacy: Using asTTle to guide teaching.* Paper presented at the New Zealand Association for the Teaching of English (NZATE) Annual Conference, Dunedin, NZ.

Andersen, I., & Brown, G. T. L. (in preparation). Flight paths through literacy: Using asTTle to guide teaching in a Year 12 alternative English class. *English in Aotearoa*.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Barnette, J., Orletsky, S., & Sattes, B. (1994, July). *Evaluation of teacher classroom questioning behaviors.* Paper presented at the Third Annual National Evaluation Institute Field Symposium, Gatlinburg, TN.

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Erlbaum.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome).* New York: Academic Press.

Biggs, J. B., & Collis, K. F. (1989). Toward a model of school-based curriculum development and assessment using the SOLO taxonomy. *Australian Journal of Education, 33*, 151-163.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead, UK: Open University Press.

Bloom, B., S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.). (1956). *Taxonomy of Educational Objectives: The classification of Educational Goals--Handbook 1 Cognitive Domain.* New York: David McKay.

Bond, L., Smith, T., Baker, W. K., & Hattie, J. A. (2000). *The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study* (Research Report). Greensboro, NC: University of North Caolina at Greensboro, Center for Educational Research and Evaluation.

Brown, G. T. L. (2002a). Student Beliefs About Learning. *Academic Exchange Quarterly*, 11-14.

Brown, G. T. L. (2002b). *Teachers' Conceptions of Assessment.* Unpublished doctoral dissertation, University of Auckland, Auckland, NZ.

Calder, J. R. (1983). In the cells of Bloom's taxonomy. *Journal of Curriculum Studies, 15*, 291-302.

Collis, K. F., & Biggs, J. B. (1986). Using the SOLO taxonomy. ***set**: Research Information for Teachers,* (2), 4.

Collis, K. F., & Davey, H. A. (1986). A technique for evaluating skills in high school science. *Journal of Research in Science Teaching, 23*, 651-663.

Collis, K. F., & Romberg, T. A. (1992). *Collis-Romberg mathematical problem solving profiles*. Melbourne, Aus: ACER.

Courtney, T. D. (1986). The significance of the SOLO taxonomy for learning and teaching in geography. *Geographical Education, 5*, 47-50.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership, 43*(2), 45-48.

Furst, E. J. (1981). Bloom's taxonomy of educational objectives for the cognitive domain: Philosophical and educational issues. *Review of Educational Research, 15*, 175-198.

Gall, M. D. (1970). The use of questions in teaching. *Review of Educational Research, 40*, 707-721.

Gall, M. D. (1984). Synthesis of research on teachers' questioning. *Educational Leadership, 42*(3), 40-47.

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.

Hamben, K. (1984). An art criticism questioning strategy within the framework of Bloom's taxonomy. *Studies in Art Education: A Journal of Issues and Research, 26*, 41-50.

Hattie, J. A., & Purdie, N. (1994). *Using the SOLO taxonomy to classify test items.* Unpublished manuscript, University of Western Australia, Graduate School of Education, Perth, Aus.

Hattie, J. A., & Purdie, N. (1998). The SOLO model: Addressing fundamental measurement issues. In B. Dart & G. Boulton-Lewis (Eds.), *Teaching and Learning in Higher Education* (pp. 145-176). Melbourne, Aus.: ACER.

Hattie, J. C., Biggs, J. B., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*(2), 99-136.

Holbrook, J. B. (1989). *Writing chemistry items using the SOLO taxonomy.* Paper presented at the Sixth Annual Conference of the Hong Kong Educational Research Association, Hong Kong.

Ki, W. W. (1989). *Computer Studies.* Paper presented at the Sixth Annual Conference of the Hong Kong Educational Research Association, Hong Kong.

Kloss, R. J. (1988). Toward asking the right questions: The beautiful, the pretty, and the messy ones. *The Clearing House, 61*, 245-248.

Kryzanowski, E. M. (1988). *Attitudes towards adolescent pregnancy and parenthood.* Unpublished doctoral dissertation, University of Alberta, Edmonton, Alta.

Maguire, T. O. (1988, December). *The use of the SOLO taxonomy for evaluating a program for gifted students.* Paper presented at the Annual Conference of the Australian Association for Research in Education, Armidale, NSW.

Marton, F., & Saljo, R. (1976). On qualitative differences in learning. I - outcome and process. *British Journal of Educational Psychology, 42*, 4-11.

Marzano, R. J. (1991). Language, the language arts, and thinking. In J. Flood, J. M. Jensen, D. Lapp & J. R. Squire (Eds.), *Handbook of Research on Teaching the English Language Arts* (pp. 559-586). New York: Macmillan.

McMillan, J. H. (2001). *Classroom assessment: Principles and practice for effective instruction* (2 ed.). Boston, MA,: Allyn & Bacon.

Meagher-Lundberg, P., & Brown, G. T. L. (2001). *Item signature study: Report on the characteristics of reading texts and items from calibration 1.* (Technical Report No. 12). Auckland, NZ: University of Auckland, Project asTTle.

Pong, W. Y. (1989). *Economics.* Paper presented at the Sixth Annual Conference of the Hong Kong Educational Research Association, Hong Kong.

Pring, R. (1971). Bloom's taxonomy: A philosophical critique. *Cambridge Journal of Education, 1*, 83-91.

Schrag, F. (1989). Are there levels of thinking? *Teachers College Record, 90*(4), 529-533.

Smith, & Hattie, J. A. (in preparation).

Snow, R. E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher, 18*, 8-14.

Stimpson, P. (1989). *Viability of the SOLO test format: The case of geography.* Paper presented at the Sixth Annual Conference of the Hong Kong Educational Research Association, Hong Kong.

Thomas, G., Holton, D., Tagg, A., & Brown, G. T. L. (2002). *Numeracy item signature study: A theoretically derived basis.* (Technical Report No. 25). Auckland, NZ: University of Auckland, Project asTTle.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom.* Buckingham, UK: Open University Press.

University of Auckland, & Ministry of Education. (2003). Assessment Tools for Teaching and Learning (Version 3). Wellington, NZ: Learning Media.

Wade, S. E., & Moje, E. B. (2000). The role of text in classroom learning. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of Reading Research Volume III* (pp. 609-627). Mahwah, NJ: LEA.

Wood, D. (1988). *How Children Think & Learn*. Oxford, UK: Blackwell.

Wood, D. (1991). Aspects of teaching and learning. In P. Light, S. Sheldon & M. Woodhead (Eds.), *Learning to Think* (pp. 97-120). London: Routledge.

Appendix A: The Problems of Bloom's Taxonomy

For the past four decades the development of most measures of cognition and achievement have been based on Bloom's taxonomy of educational objectives (Bloom, Engelhart, Furst, et al., 1956). The taxonomy was published in 1956, has sold over a million copies, has been translated into several languages, and has been cited thousands of times. The Bloom taxonomy has been extensively used in teacher education to suggest learning and teaching strategies, has formed the basis of many tests developed by teachers (at least while they were in teacher training), and has been used to evaluate many tests. It is thus remarkable that the taxonomy has been subject to little research or evaluation. Most of the evaluations are philosophical treatises noting, among other criticisms, that there is no evidence for the invariance of these stages, or claiming that the taxonomy is not based on any known theory of learning or teaching (Calder, 1983; Furst, 1981).

Despite the popularity of the Bloom cognitive taxonomy, there is little support for the use of it in organising instruction, curriculum, or assessment. The fundamental difficulty in using Bloom's Taxonomy to guide the development of questioning is the false assumption that these categories represent a hierarchically ordered set (McMillan, 2001). Clearly, very young children are capable of evaluating (e.g., "that's yucky!") while they may be relatively incapable of analysing and communicating their reasons for such negative evaluations.

A new edition of Bloom's Taxonomy (Anderson & Krathwohl, 2001) recognises the limitations of the earlier (1956) version, changes the order of the cognitive processes hierarchy (remember, understand, apply, analyse, evaluate, and create), and introduces a new dimension of knowledge types (i.e., factual, conceptual, procedural, and metacognitive). Thus, a matrix of 24 process-object cells is described, wherein each cognitive process has four types of knowledge. For example, the remembering process can include factual knowledge (i.e., knowledge of terminology or details), conceptual knowledge (i.e., knowledge of categories, principles, theories), procedural knowledge (i.e., knowledge of skills, methods, techniques), or metacognitive knowledge (i.e., knowledge of strategy, context, conditions). With these two axes, the revised Bloom taxonomy describes both the intended cognitive process and the category of knowledge underpinning an educational objective, outcome, or task. It is asserted that the two dimensions are arranged in hierarchical order of increasing cognitive complexity. Metacognitive knowledge is more abstract than the concrete factual knowledge and creative cognition is more complex than remembering. Whatever the merits of this reincarnation of the Bloom model, the most important addition is that it movement to a surface to deep continuum.

There are numerous difficulties with the Bloom Taxonomy (1956 or 2001), and any competing model needs to ensure it does not have similar difficulties. Hattie and Purdie (1998) detailed these problems and they included:

- The Bloom taxonomy presupposes that there is a necessary relationship between the questions asked and the responses to be elicited (see Schrag, 1989), whereas in the SOLO taxonomy both the questions and the answers can be at differing levels.
- Bloom separates 'knowledge' from the intellectual abilities or process that operate on this 'knowledge' (Furst, 1981), whereas the SOLO taxonomy is primarily based on the processes of understanding used by the students when answering the prompts. Knowledge, therefore, permeates across all levels of the SOLO taxonomy.
- Bloom has argued that his taxonomy is related not only to complexity but also to an order of difficulty such that problems requiring behaviour at one level should be answered more correctly before tackling problems requiring behaviour at a higher level. Although there may be measurement advantages to this increasing difficulty, this is not a necessary requirement of the SOLO method. It is possible for an item at the relational level, for example, to be constructed so that it is less difficult than an item at the unistructural level. For example, an item aiming to elicit relational responses might be 'How does the movement of the Earth relative to the sun define day and night'. This may be easier (depending on instruction, etc.) than a unistructural item that asks 'What does celestial rotation mean?' Thus, there can be certain aspects of knowledge that are more complex than aspects of analyses or evaluation (see also Furst, 1981; Pring, 1971).

- Bloom's taxonomy is not accompanied by criteria for judging the outcome of the activity (Ennis, 1985), whereas SOLO is explicitly useful for judging the outcomes. Take for example, a series of art questions suggested by Hamben (1984) (Figure 10). When using Bloom's taxonomy, the supposition is that the question leads to the particular type of Bloom response. There is no necessary relationship, however, as a student may respond with a very deep response to the supposedly lower order question: 'Describe the subject matter of Guernica?' Similarly, a student may provide a very surface response to 'What is your opinion of Picasso's Guernica'?

Figure 10
Bloom's Taxonomy Applied to Questions/Tasks about Picasso's *Guernica*

<u>Knowledge</u>. Who painted *Guernica*?

<u>Comprehension</u>. Describe the subject matter of *Guernica*.

<u>Application</u>. Relate the theme of *Guernica* to a current event.

<u>Analysis</u>. What compositional principles did Picasso use in *Guernica*?

<u>Synthesis</u>. Imagine yourself as one of the figures in *Guernica* and describe your life history.

<u>Evaluation</u>. What is your opinion of Picasso's *Guernica*?

The greatest criticism of the Bloom taxonomy is that there is little evidence supporting the invariance and hierarchical nature of the six levels. Bloom claimed these six levels "represent something of a hierarchical order of the different classes of objectives. As we have defined them, the objectives in one class are likely to make use of and be built on the behaviors found in the preceding classes in this list" (1956, p. 18). A prior condition of the hierarchy is that there is common understanding of the various levels. Ennis (1985) argued that analysis relates to many levels. "Analysis of a chemical compound, analysis of an argument, analysis of a word, analysis of an opponent's weaknesses in a basketball game, and analysis of the political situation in South Africa seem like such different activities that we might very well wonder just what we are supposed to teach under the label 'analysis' “(p. 45). Calder (1983) was much more critical of all Bloom’s levels, and provided illustration of the conceptual morass that followed from a 'classification filled with nebulous terms (which) makes it impossible to detect similarities in objectives in different subject areas, and frustrates efforts to develop precise principles of teaching and testing bearing on sharply delineated objectives' (p. 297). As an example, he considered the notion of 'knowledge', which includes cases where the student relates definitions of terms to specific instances as well as recalls definitions verbatim. This first instance could be confused with 'relating abstractions to concrete instances' which is defined by Bloom as Comprehension. Further, he claimed that “too many categories contain a pantechnicon assortment of skills” (p. 298). Such conceptual confusions make it difficult to expect there to be a hierarchy.

Appendix B. SOLO-Based Teacher Observation Questionnaire

Directions: Please consider the lesson that has just been taught and respond to each question by circling ONE of the responses.

Strongly Disagree	Disagree	Somewhat Agree	Agree	Strongly Agree
1	2	3	4	5

Capacity	**SD**	**D**	**SA**	**A**	**SA**
1. The students had to think *hard* in the lesson	1	2	3	4	5
2. During the lesson, the students had to concentrate on one idea at a time	1	2	3	4	5
3. In this lesson the students had to concentrate on a few ideas at a time	1	2	3	4	5
4. During the lesson, the students had to concentrate on many ideas at a time	1	2	3	4	5
5. In this lesson, the students were asked to focus on recalling information from previous lessons	1	2	3	4	5
6. In the lesson, the students had to relate many ideas, or extend beyond the ideas presented	1	2	3	4	5

Consistency and Closure	**SD**	**D**	**SA**	**A**	**SA**
7. A major aim of this lesson was for students to experience a sense of completion/conclusion/closure before they moved onto the next lesson	1	2	3	4	5
8. In this lesson, an aim was that students would end the lesson with unresolved questions and/or contradictions	1	2	3	4	5
9. The students had to think about open-ended issues or problems either during or at the end of the lesson	1	2	3	4	5
10. During this lesson a major goal was for the students to place facts into a given structure	1	2	3	4	5

Structure and Capacity	**SD**	**D**	**SA**	**A**	**SA**
11. In this lesson, students had to learn many facts, details or concepts	1	2	3	4	5
12. A major goal of this lesson was for students to reproduce basic facts, details, or concepts	1	2	3	4	5
13. A major goal of this lesson was for students to synthesize or integrate concepts	1	2	3	4	5
14. A major goal of today's lesson was to have students make use of a concept or principle underlying the content or task	1	2	3	4	5

Relationship	**SD**	**D**	**SA**	**A**	**SA**
15. Students had to find relationships between ideas during the lesson	1	2	3	4	5
16. During this lesson, students had to find relationships between a few ideas or facts	1	2	3	4	5

17. During this lesson, a major goal was for students to construct generalizations to broader or novel ideas/situations 1 2 3 4 5
18. A major goal was to have students extend their understanding beyond the content of today's lesson 1 2 3 4 5

Level of Thinking

Please use the following scale:

Not at all	*Occasionally*	*About half the time*	*Frequently*	*Nearly all the time*
1	*2*	*3*	*4*	*5*

19. **Prestructural**. There may have been preliminary preparation, but the task itself is not attacked in an appropriate way. 1 2 3 4 5
20. **Unistructural**. One aspect of the task is performed or understood serially. However, there is no relationship to other factors or ideas. 1 2 3 4 5
21. **Multistructural**. Two or more aspects of the task are performed or understood serially with limited inter-relationships to other ideas. The individual has limited understanding of how concepts or ideas fit together. 1 2 3 4 5
22. **Relational**. Several aspects are integrated so that the whole has a coherent structure and meaning in and of itself. It is like fine woven fabric. 1 2 3 4 5
23. **Extended Abstract**. The coherent whole is extended to abstract principles of generalizations underlying what is being taught. 1 2 3 4 5

Based on the evidence you have just seen,

23. Are the teacher aims and the teaching in the class lesson representative of surface or deep levels of understanding *Surface* *SomewhatSurface* *SomewhatDeep* *Deep*

24 Please provide a brief statement of the evidence available for the surface to deep level responses in Question 23

25. Are student outcomes in the class lesson representative of surface or deep levels of understanding? *Surface* *Somewhat Surface* *Somewhat Deep* *Deep*

26. Please provide a brief statement of the evidence available for the surface to deep level responses in Question 25

27. Note that your judgement above essentially classifies the students' outcomes as either surface or deep. In your judgement, which SOLO level below best describes the preponderance of the student achievement outcomes evidence?

    a. Unistructural
    b. Multistructural
    c. Relational
    d. Extended Abstract