Assessing Chinese Students' Understandings in Mathematics and Science: Ensuring Valid and Reliable Data within International Contexts

Chunxia Wang, Dr Debra Panizzon & Professor John Pegg

Abstract

Assessment practices in China are based largely upon the quantification of students' understandings, which are measured using multiple-choice tests based upon rote-learning and memorisation. These methods differ from global trends in education where there is a move towards assessment approaches that attempt to identify what students 'know' and 'can do'. To explore alternative modes of assessing the conceptual understanding of Chinese students, a study was designed based upon the Structure of the Observed Learning Outcome (SOLO) model as the theoretical framework. While the model has been used extensively in many Western countries including Australia, United States and the United Kingdom, this was the first study to be undertaken with Year 5 students in China. This paper describes the research issues that emerged from the pilot study including problems in ensuring the accuracy of translations from Chinese to English, the need to provide Chinese students with the opportunity to provide their 'best possible' explanations, and the importance of the researcher being present for the collection of data. Once identified, each of these issues was considered carefully in the design and implementation of the main study to ensure improved validity. With this achieved, emerging patterns demonstrate the applicability of the SOLO model not only for assessing student understanding but in helping to identify a learning trajectory for mathematics and science concepts for these Chinese students.

In the last few years there has been a global trend towards outcomes-based education. This change in focus requires a move from assessment as a way of measuring and quantifying student achievement, to one that gauges the depth and *quality* of students' understandings (Bell & Cowie 2001; Black & William 1998; Treagust et al. 2001). One of the major challenges for teachers has been the need to *embed* assessment into the teaching and learning process so that it is used to provide constructive feedback to students and teachers (Black & Harrison 2002; Wilson & Sloane 2000). In this new climate, teachers consider both the purpose and types of assessment used in their classrooms and how these tasks provide information on the 'level' of student's understanding. The implications of teachers having this information is that it can lead to more focused teaching that enhances student learning in discipline areas (Biggs 1996). While this has been a difficult transition for educational systems in western countries (Pegg & Panizzon 2006), it represents major challenges in countries where there is a strong tradition towards examinations.

China presents just such a context. From the beginning of civil service examinations (A.D. 581-617) to the year they were abandoned in 1905, examination success and civil service careers were inextricably linked. Together, they provided access to wealth and membership of the ‘upper’ class in China (Pepper 1996). Even today, this examination-performance-driven meritocracy can still be seen as an important motivating force for learning and career achievement irrespective of social class in China (Delors 1996).

In this paper we discuss the pilot study undertaken as part of a larger study to investigate implications associated with assessing Chinese students’ understandings of science and mathematics concepts using the SOLO model. Initially, we outline the educational context of the study including an overview of the SOLO model. The purpose and design of the pilot are then detailed followed by a summary of the research findings. Finally, we discuss the research issues that emerged, and how these were considered in the main study to ensure the validity of the results.

Chinese Educational Context

China has always valued education highly recognising it as a means of ensuring the success of individuals and society as a ‘whole'. In order to understand the struggle, experience, and emphasis of assessment in the current Chinese education system, it is necessary to provide an historical glimpse of the complex relationship between traditional Confucian educational theories and the long history of the civil examination system.
The importance of education in Chinese culture is derived from the teaching of Confucius. In theory, the end-product of Confucian education was the cultivation of a person in the ethical sense. In practice, however, it was concerned more with the preparation and selection of a ruling elite (Pepper 1996). The classic text underpinning education was the Confucian Canon comprising a collection of important writings that have been studied and examined over the past 2,000 years. These texts provided the context of Confucian teaching and learning, and are usually referred to as the five classics (*Wu Jing*) and four books (*Si Shu*).

The beginning of civil service examinations in China occurred during the Sui Dynasty (A.D. 581-617) and lasted for 1300 years. These examinations were abandoned in 1905. To pass the different levels of examinations from the Sub-Prefect, Prefect and Ju Ren (Master’s degree) (Purcell 1936), the selection procedures required strict assessment standards supported by a stringent method of implementation. Candidates for these examinations were expected to memorise the five classics and four books, then reproduce the writings. The emphasis was on memorising huge amounts of material from the classics so as to write highly structured essays. All knowledge resided in classical writings, called the ‘eight-legs’ (Ba Gu Wen), a kind of full-dress formal essay.
Essentially, the education system was about selecting the most talented individuals in the country for the service of the State (Tsang 1968). During this long period, Confucian learning, imperial power, and

bureaucratic authority were bound together. Taken in total, these aspects extorted significant influence over Chinese life and society (Pepper 1996).

Western learning ideas and teaching methods did not emerge in China until the western intrusion into China at the end of the 19th century (Pepper 1996). As a consequence of this 'melding of minds', China began considering changes in education at the beginning of the 20th century. During the 1920s, the content and objectives of examinations were extended, a change influenced directly by the United States and European educational systems (Han & Yang 2001). During this period, Professor Liao Shicheng and Chen Heqin collaborated on a book titled *Methods of Intelligence Testing* (Liao & Chen 1921). This was the first time Western psychometrics was introduced into China by Chinese academics.

Although the imperial examination system was abolished almost 100 years ago, its impact is still evident in China today with examinations considered the most important component of education in China. The examination focus has influenced the purpose and curriculum content of Chinese education and has resulted in particular teaching methods and learning strategies in the classroom (Lee 1985). In particular, memorisation and rote-learning strategies underpin the function of education, overshadowing the implementation of creative thinking and problem-solving approaches. As a consequence, there are high parental and societal expectations, exerting great pressure on Chinese youth to excel in examinations.

In recent years, the examination-driven educational focus has aroused debate and controversy among Chinese educators with a view that China must embrace the move towards an outcomes-based approach to education. However, this is difficult given the historical traditions and the need to change the 'culture' in terms of the purpose and importance of education generally (Yang 2002). From educational decision-makers to school teachers, there is a call to reform the current educational assessment practices from focusing on how *much* students know to *how well* it is understood and can be transferred to other learning situations. In keeping with this view, Xian (2005) suggested that teachers should be encouraging students to provide greater explanations and opinions, along with opportunities to demonstrate analysis, reasoning, and discussion skills. This view contrasts greatly to the traditional rote-learning and fact-listing strategies required to achieve success in current examinations.

Clearly, educational reform is underway to improve the quality of learning and teaching so as to ensure a positive future for Chinese students. This reform is aimed at fostering innovative capabilities and improving the practical competence of all students (Jiang 2002; Yang 2002). Central to this idea of change in education is assessment. In particular, assessment should be organised so that it provides information to the teacher and student about the quality of understanding and where teaching might most profitably be directed.

Assessing the Quality of Students' Understandings

The work of Biggs and Collis (1982, 1991) has provided a powerful model to assist the evaluation of the quality of learning outcomes. The SOLO model as a framework offers a systematic way of observing what appears to be a consistent sequence or pattern of learning (learning cycle) displayed by learners within a variety of school-based tasks (Biggs & Collis 1989). Underlying the model is the assumption that cognitive understanding does not equate to a stable cognitive construct as with Piaget (1954), but involves individual characteristics that are content and context dependent. Over the past 20 years, the SOLO model (Biggs & Collis 1982, 1991) has been used in a variety of subject and topic areas, including number and operations, history, geography, poetry, language learning, science and history (Pegg 2003). The research and practice of SOLO has been carried out in Australia, USA, Hong Kong, UK and many other countries.

Incorporated in the SOLO model are two important features. The first concerns the nature or abstractness of the response and is referred to as the *mode of thinking*. It refers to the type of intellectual functioning required to address adequately a particular stimulus, e.g. mathematics question. The second feature depends on an individual's ability to handle, with increased sophistication, relevant cues. This feature is referred to as *levels of response*, which are seen to reside within cycles of learning that provide a hierarchical description of the nature of the structure of a response. While these levels occur within each mode, the specific nature of these levels is dependent on the particular mode targeted by the stimulus item.

The five *modes of thinking* are:

- *Sensorimotor* where a person reacts to the physical environment. For the very young child it is the mode in which motor skills are acquired. In adult life, the mode is utilised as skills associated with various sports develop and evolve.
- *Ikonic* when a person is able to internalise actions in the form of images. It is in this mode that the young child develops words and images that represent objects and events. For the adult, this mode of functioning assists in the appreciation of art and music and leads to a form of knowledge referred to as intuitive.
- *Concrete-symbolic* when a person thinks by the use of a symbol system such as written language and number systems. Thinking in this mode is reliant on a 'real-world' referent. This is the most common mode addressed in learning in the upper primary and secondary school.
- *Formal* where a person is able to consider concepts that are abstract. Students are no longer restricted to a real-world referent and are able to work with 'principles' and 'theories.' In its more advanced form it involves the development of disciplines.

- *Post Formal* when a person is able to question or challenge the fundamental structure of theories or disciplines (Panizzon 2003).

While the five modes of thinking are distinct, the functioning in a later acquired mode (say, concrete-symbolic) does not preclude the use of an earlier acquired mode to support understanding (such as ikonic or sensorimotor). This is referred to as multi-modal functioning and supports other theories, such as Gardner's Multiple Intelligences (Gardner & Hatch 1989). The important consideration here is that once a mode has been accessed, then all earlier acquired modes are available and continue to develop throughout life in response to experiential, social, cultural, educational, or genetic factors (Collis et al. 1998).

In terms of the second feature of the model, three *levels of responses* comprise a cycle of learning within a particular mode. These are referred to as:

- *Unistructural* (U): where the individual focuses on the domain/problem, but uses only one piece of relevant data so the response may be inconsistent.
- *Multistructural* (M): where two or more pieces of data are identified as independent units. No integration is demonstrated between the data with inconsistencies often evident in the response.
- *Relational* (R): where all data are now available, with each piece woven into an overall mosaic of relationships culminating in a logical endpoint. The whole has become a coherent structure lacking inconsistencies within the given context.

The modes of thinking and levels of responses within these modes are summarised in Figure 1.

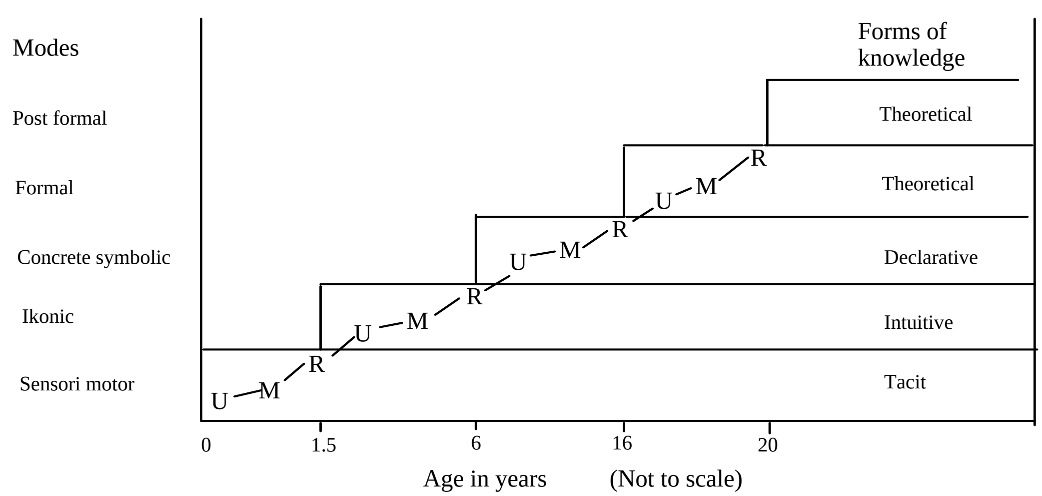

Figure 1: *Modes and levels in SOLO Model (Adapted from Biggs & Collis 1991)*

While the U, M, and R levels identify a cycle of learning in a mode, research studies over the few years have indicated that a single cycle is restrictive and does not explain fully the development of a concept (Campbell et al. 1992; Collis et al. 1998; Levins & Pegg 1993; Panizzon 2003; Panizzon et al. 2005; Pegg 1992, 2003; Watson et al. 1995). Subsequently, two unistructural – multistructural – relational cycles (i.e., U_1-M_1-R_1 and U_2-M_2-R_2) have been incorporated into the model for the concrete-symbolic and formal modes (see Figure 2).

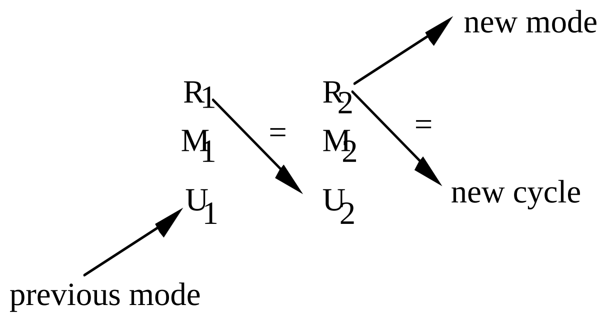

Figure 2: *Two learning cycles in the concrete-symbolic mode*

In Figure 2 the 'previous mode' refers to the ikonic mode and the six levels represent two cycles of learning in the concrete-symbolic mode. The diagram illustrates the general development pattern expected moving from U_1 to M_1 to R_1 etc. The equal sign on the pathway from R_1 to U_2 indicates diagrammatically that for many students this development is more about mental chunking of the ideas as opposed to the addition of necessarily new content. The 'new cycle' represents potentially a further cycle of development in the concrete-symbolic mode and the 'new mode' in this example refers to the formal mode.

Research Method

The overall aim of the study discussed in this paper is to identify developmental pathways in students' conceptions of six different concepts (three in mathematics and three in science). In addition, interest was directed at exploring the differences and similarities in student's cognitive growth both within and across science and mathematics. Information of this type would provide a basis for teachers and other educators in assessing students' understandings and in selecting particular teaching strategies to enhance students' learning. However, given that the SOLO model had not been used in China previously, the purpose of the pilot study was to investigate the appropriateness of the SOLO model (Biggs & Collis 1982, 1991) as a framework and identify aspects of the research design necessary to provide valid results.

Curriculum Context

Mathematics and science subjects are compulsory for students in Years 1-9. Both subjects in Year 5 use the National Standard Curriculum of China. Summative assessment tasks are conducted in both courses and include term tests, academic year tests and entrance tests for a higher grade. Local education departments develop these tests. In addition, schools implement other assessments, such as tests before the delivery of a new topic, at the completion of a topic, and after an extended period time. Clearly, formal testing is a major form of assessment in schools in China.

Sample

Contact was made with a Chinese school located in Wu Han, the capital city of Hu Bei Province that had been visited previously by the researcher. With approval received from the principal, approximately 50 Year 5 students were invited to complete two questions providing their responses in Chinese. The sample consisted of an even distribution of female and male students.

Pilot Study Design

Guiding the study were:

Theme 1: Identify developmental pathways in students' understanding in mathematics and science.

Theme 2: Explore underlying thinking traits in Year 5 students undertaking mathematics and science.

To identify relevant topic areas for the Chinese students, relevant curriculum materials were accessed and explored. From these documents two items (one in science, one in mathematics) were selected. Both these items had been used with Australian primary students in Years 5 and 6 thereby supporting the construct validity of the items. Subsequently, it was possible to explore the similarities and differences that emerged between the two groups of students. The two questions were:

i. Explain how you would divide nine (9) round shape birthday cakes between sixteen (16) people so that each person receives the same amount? (mathematics)

ii. Describe it as much detail as possible why we get day and night on the Earth? (science)

These questions were written in Chinese and posted to the Year 5 teachers in the school. Teachers then wrote the questions on the board during a lesson allocated to the task. All the participants answered the questions in Chinese while the teachers were present. The students were encouraged to provide as much detail in their explanations as they could while addressing each question. Once completed, the responses were posted back to Australia where they were translated into English by the researcher, and then analysed using the SOLO model.

Data analysis

The first critical task was to translate student responses from Chinese into English for coding purposes thereby enabling two experienced researchers with SOLO to verify the accuracy of the coding. The English translation was a critical component of the research and had to be undertaken sensitively and rigorously to ensure the reliability of translation (De Vijver & Leung 1997; Twinn 1997). Therefore, a native Chinese speaker who was an academic at the University of New England checked each translated script for accuracy and met with the researcher to cross-check the procedure.

With translation completed, the researcher and co-researchers (experts with the SOLO model) sorted responses into ‘like’ categories. Once established, each category was considered in light of the SOLO model identifying the mode of thinking and level of understanding being demonstrated. The reliability of the categorisation of responses was checked using independent rankings from the two colleagues with an overall agreement percentage calculated (i.e. 93-96%). This was undertaken for the two questions by comparing the codings between two researchers (Denzin 1997; Wiersma 1991).

Findings from Pilot Study

The initial findings are presented in relation to the research themes explored in the pilot study and the important aspects about research design that emerged from the process. The analysis of students’ responses provided two cycles of learning in the concrete-symbolic mode in both mathematics and science subjects. Although a detailed comparison of the responses was not a part of the pilot study, the responses provided by the students were similar in structure to those of Australian students of the same age. So, despite the differences in culture and language, the scripts from two different countries suggest that students demonstrate a consistent hierarchical pathway in learning these concepts.

Students’ understandings of fractions

The responses displayed an increasing degree of complexity. Student responses that were coded in the first learning cycle of the concrete-symbolic mode (U_1, M_1, R_1) used diagrams and written cues to address the question. They did not demonstrate an understanding of the nature of the question relating to notation and the operations involved. Hence, there was no attempt to solve the problem using mathematical notation. In other words, most of the students started dealing with the problem in such a way that most responses were focused on the action of dividing nine cakes rather than the effect of the action. Therefore, the most common difficulty to emerge for students working in this cycle was how to allow each person to receive the same amount of cake by the division.

In the second learning cycle of the concrete-symbolic mode (U_2, M_2, R_2), students attempted to use fraction notation and operations in their solutions. The best responses had the students expressing fractions in terms of common denominators and completing a simple addition. This was coded as M_2 in the concrete-symbolic mode.

The following chart provides an overview for the type of responses provided for the development of understanding about the concept of a fraction. The responses within the first cycle of the concrete-symbolic mode describe the actions and consequences (see Figure 3).

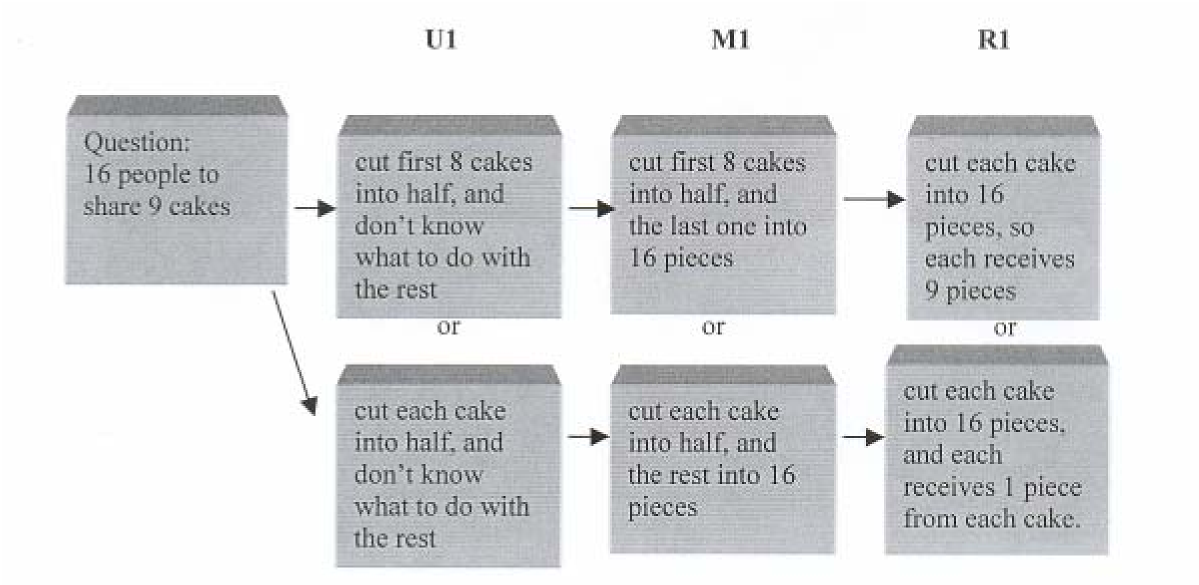

Figure 3: *Examples of student responses for the first learning cycle in the concrete-symbolic mode*

The responses within the second cycle use fraction notation to describe the effect of their actions (see Figure 4).

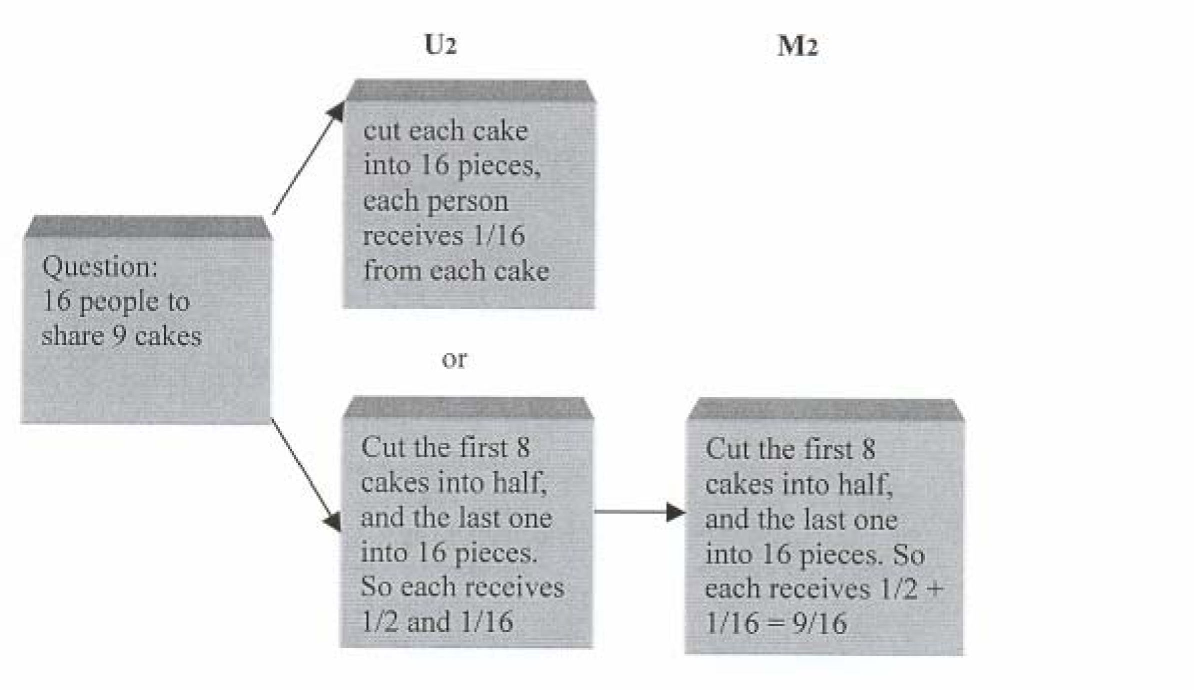

Figure 4: *Examples of student responses for the second learning cycle in the concrete-symbolic mode*

Students' understandings of day and night

The coding of all the responses for the concept of day and night demonstrated that the majority of Chinese students in Year 5 were capable of expressing and reasoning a scientific concept within the concrete-symbolic mode. As with the fractions question, two cycles of learning within the concrete-symbolic mode were identifiable.

Responses within the first cycle (U_1, M_1, R_1) indicated that students were aware that the change of day and night had something to do with the position or movement of the Earth and the Moon in relation to the Sun. While students recognised that daylight meant that one side of the Earth received day and the other side night, their explanations of how and why this occurred were limited. Interestingly, a number of alternative conceptions emerged in relation to 'the Sun rising and falling' and ideas about the 'moon only being visible at night':

> *When the Earth moves towards the Moon, night is coming, because the Moon is dark. When the Earth moves towards the Sun, the day is coming (U_1 with focus on movement of the Earth).*
> *The side of the Earth facing the Sun is day, otherwise it is night. The Earth is round, it moves around the Sun. The light from the Sun reflects on the Earth, that is how we get day and night (M_1 with focus on movement of the Earth, the position of the Earth and the Sun, and descriptions about the behaviour of these bodies).*

> *How does the Earth get day and night? I think it is because the Earth spins all the time. The Sun doesn't move at all. It takes 24 hours for the Earth to spin one cycle. The Earth gets day is because one side of the Earth is facing the Sun, and the other side is opposite the Sun, so it gets night (R_1 with the interrelationships between factors identified and a logical endpoint attained).*

In the second cycle of the concrete-symbolic mode (U2, M2, R2), explanations focused on the spinning of the Earth on its axis. There was recognition that it was this spinning that caused light on one side of the Earth and darkness on the other. Some students were then able to use this information to explain that the length of daylight and night varied depending on where one was positioned on the Earth. Therefore, explanations in this cycle demonstrated a scientific description of one or more factors involved in the phenomena and how these factors interrelate to each other:

> *The Earth is moving, and the Moon and the Sun are moving too. The Sun is the centre of the Solar system. The Earth spins on the axis, if it doesn't meet the Moon at the same spot, the Earth gets day. Otherwise, the Moon will shadow the Earth, so the Earth gets night (U_2 with realisation that the axis is critical to day and night).*
>
> *The Earth spins on its axis from West to East while it orbits around the Sun from west to east. The Earth is spinning causes the change of the day and night, its orbiting around the Sun brings the change of the four seasons (M_2 with the axis identified as important, but also direction of spin although an alternative conception in this example).*

Clearly, there is a difference in thinking of students between the first and second cycles of the concrete-symbolic mode. In the case of the first cycle, students described planetary movement generally while those in the second cycle demonstrated some degree of explaining the process behind day and night. Again, these results were similar to those identified for Year 5 and 6 students in Australia. This adds further support to suggesting the applicability of the SOLO model for Chinese students.

Methodological Considerations Emerging

A number of operational and technical issues emerged during the pilot study that required careful consideration in developing the design for the main study. These included the:

- Use of questions that were contextually relevant to the students
- Importance of accuracy in the translation process
- Need for a larger student sample to obtain a variety of responses
- Opportunity for students to provide their 'best possible' responses
- Need to ensure a degree of consistency in how the data were collected.

First, in relation to the questions used with students, the fraction item raised concerns because some of the students did not understand ‘a cake’ as this was not a commonly used term in Chinese culture. Although the Year 5 teacher (in China) had checked the question prior to its use, students raised questions that suggested that the concept was ambiguous. While cakes of various shapes and sizes were relevant for the majority of students, this was clearly not the case across the board and suggested that an alternative object should be used in further research. Subsequently, this question was altered for the main study.

Second, the fundamental aspect underpinning the validity of the data was the process used to ensure the accuracy of translation. As the process used for the pilot appeared successful, it was adopted for the main study. Initially, the researcher translated all written responses to each question. During a meeting with a Chinese-speaking senior academic from UNE, all of the responses for each of the questions were checked for accuracy.

Third, the pilot study identified that selection of students from one or two classes in Year 5 did not provide a range of responses representative of the various levels of the SOLO model. To address this issue, the entire cohort of Year 5 students was invited to complete the questionnaires in the main study. This resulted in 150 participants answering the questions.

Fourth, additional probing of students about particular statements made in their explanations would improve the quality of the data obtained. Consequently, interviews became a feature of the main study as they provided an opportunity to explore students’ understandings at a deeper level, clarify student explanations, and also allow any changes in thinking to be identified. Therefore, approximately 25% of students in the school Year 5 cohort were interviewed using purposive sampling.

Finally, it was decided that the researcher should collect the data in China. Most importantly, her presence would ensure consistency in the information provided to students and the questionnaire process. Furthermore, it would allow her to conduct the interviews in person, hence improve the depth of data obtained. The other major advantage was that it allowed the researcher to gain a greater awareness of the educational context while allowing observations of classrooms and interviewing of teachers about their views of teaching and learning to be undertaken. Although not incorporated into the final study, these insights provided an important background in interpreting the results at a broader level.

Overall, these findings demonstrate clearly the necessity of undertaking a pilot study as a means of trialling and fine-tuning research techniques thereby ensuring the potential for greater reliability and validity of the data. In planning for the main study, each of these issues was addressed resulting in a well-structured design that minimised the impact of potential ‘threats’ to the study. In addition, it ensured that the researcher was suitably confident in knowing *what needed to be undertaken* and *how this should be achieved* in an ethical manner.

Conclusion

The study outlined in this paper and the findings generated demonstrate the applicability of the SOLO model in the Chinese educational context. As with the Australian students, two learning cycles were identifiable in the concrete-symbolic mode for both questions. However, equally important were the methodological issues that emerged as the researcher worked through the data set. Careful consideration of these factors at this stage ensured a sound and robust design for the main study thereby maximising the validity of the final results. Unexpectedly, however, the pilot study identified a number of logistical requirements that could have hampered the research if not considered carefully prior to commencing the main study in a foreign country. This is an important outcome given that many international students work away from their supervisors during the data collection phase with minimal contact available. A staged research design that is carefully planned helps ensure that most of the major difficulties are dealt with prior to the researcher's departure.

REFERENCES

Bell, B. & Cowie, B. 2001, 'The characteristics of formative assessment in science education', *Science Education,* vol.85, pp.536-553.

Biggs, J. 1996, 'Enhancing teaching through constructive alignment', *Higher Education*, vol.32, pp.347-364.

Biggs, J. & Collis, K. 1982, *Evaluating the Quality of Learning: The SOLO Taxonomy*, Academic Press, New York.

Biggs, J. & Collis, K.F. 1989, 'Towards a model of school-based curriculum development and assessment: Using the SOLO Taxonomy', *Australian Journal of Education,* vol.33, no.2*,* pp.151-163.

Biggs, J. & Collis, K. 1991, Multimodel learning and the quality of intelligent behaviour, in *Intelligence, Reconceptualization and Measurement,* ed H. Rowe, Laurence Erlbaum Assoc., New Jersey, pp.57-76.

Black, P.& Harrison, C. 2002, 'Assessment for learning in science classrooms', Paper presented at the *National Association for Research in Science Teaching*, April in New Orleans.

Black, P.& William, D. 1998, 'Assessment and classroom learning', *Assessment in Education*, vol.5, no.1, pp.7-74.

Campbell, K., Watson, J. & Collis, K. 1992, 'Volume measurement and intellectual development', *Journal of Structural Learning and Intelligent Systems,* vol.11, pp.279-298.

Collis, K., Jones, B., Sprod, T., Watson, J. & Fraser, S. 1998, 'Mapping development in student's understanding of vision using a cognitive structural model, *International Journal of Science Education,* vol.20, no.1, pp.44-66.

Delors, J. 1996, *Learning: The Treasure Within*, Report to UNESCO of the International Commission on Education for the 21st Century, UNESCO Publishing.

Denzin, N. K. 1997, Interpretive Ethnography: Ethnographic Practices for the 21st Century, Sage, Thousand Oaks, California.

De Vijver, F.J.R. & Leung, K. 1997, *Methods and Data Analysis for Cross-Cultural Research,* Sage Publications Inc., Thousand Oaks, California.

Gardner, H. & Hatch, T. 1989, 'Multiple intelligences go to school: Educational implications of the theory of multiple intelligences', *Educational Researcher*, vol.18, no.8, pp.4-9.

Han, M. & Yang, X. 2001, 'Educational assessment in China: Lessons from history and future prospects', *Assessment in Education*, vol.8, no.1, pp.5-10.

Jiang, S. 2002, 'The overview of study learning in middle and high schools in Chinese', *Education Practice & Research*, vol.2, pp.12-13.

Lee, T. H. C. 1985, *Government Education and Examinations in Sung China*, Chinese University Press, Hong Kong.

Levins, L. & Pegg, J. 1993, 'Students' understanding of concepts related to plant growth', *Research in Science Education,* vol.23, pp.165-173.

Liao, S. & Chen, H. 1921, *Method of Intelligence Testing,* Business Printing House, Shanghai.

Panizzon, D. 2003, 'Using a cognitive structural model to provide new insights into students' understandings of diffusion', *International Journal of Science* Education, vol.25, no.12, pp.1427-1450.

Panizzon, D., Pegg, J. & McGee, S. 2005, Incorporating different assessment tasks to gauge student understandings of planetary processes. Paper presented at the *Annual Conference for the Australian Association for Research in Education*, Melbourne, 28th November-2nd December, 2004, pp.1-18. Retrieved from
http://www.aare.edu.au/04pap/abs04.htm

Pegg, J. 1992, 'Assessing students' understanding at the primary and secondary level in the mathematical sciences', in *Reshaping Assessment Practice: Assessment in the Mathematical Sciences under Challenge,* eds J. Izard & M. Stephens, Australian Council of Educational Research, Melbourne, pp.368-385.

Pegg, J. 2003, 'Assessment in mathematics: A developmental approach', in *Mathematical Cognition,* ed M. Royer, Information Age Publishing, Greenwich, Connecticut.

Pegg, J. & Panizzon, D. 2006, Addressing changing assessment agendas: Experiences of high school mathematics teachers, Paper presented for the *American Educational Research Association Conference*, San Francisco, California, 7-11th April.

Pepper, S. 1996, Radicalism and Education Reform in 20th-century China, Cambridge University Press, UK.

Piaget, J. 1954, *The Construction of Reality in the Child,* Basic Books, New York.

Purcell, V. 1936, *Problems of Chinese Education,* Kegan Paul, Trench, Trubner & Co. Ltd, London.

Treagust, D.F., Jacobowitz, R., Gallagher, J.L. & Parker, J. 2001, 'Using assessment as a guide in teaching for understanding: A case study of a middle school science class learning about sound', *Science Education,* vol.85, pp.137-157.

Tsang, C. 1968, *Society, Schools and Progress in China*, Pergamon Press, London.

Twinn, S. 1997, 'Exploratory study examining the influence of translation on validity and reliability of qualitative data in nursing research', *Journal of Advanced Nursing,* vol.26, no.2, pp.418-423.

Watson, J., Collis, K., Callingham, R. & Moritz, J. 1995, 'A Model for Assessing Higher Order Thinking in Statistics', *Educational Research and Evaluation,* vol.1, no.3, pp.247-275.

Wiersma, W. 1991, *Research Methods in Education: An Introduction,* Allyn and Bacon Publishers, Boston.

Wilson, M. & Sloane, K. 2000, 'From principles to practice: an embedded assessment system', *Applied Measurement in Education,* vol.13, no.2, pp.181-208.

Xian, C. H.. 2005, 'Quality education does not contradict with exams', *Chinese Education Newspaper*, dated 26 December 2005.

Yang, Z. 2002, 'The difficulties and coping strategies for developing and implementing study learning curricula in Chinese', *Education Practice and Research*, vol.2, pp.16-18.