International Journal of Learning, Teaching and Educational Research
Vol. 7, No. 1, pp. 201-216, August 2014

An Evidence-based Approach to the Design of a Learning Program: Evaluating Preliminary Data Sets

Andrea Carr
University of Tasmania
Hobart, Australia

Jo-Anne Kelder
University of Tasmania
Hobart, Australia

Juliet Sondermeyer
University of Tasmania
Hobart, Australia

Abstract. The Global Perspectives (GP) program is an evidence-based curriculum initiative that integrates the process of designing and implementing a learning program with a process for developing and implementing a plan to evaluate it for effectiveness and impact. The GP program educational evaluation and research (EER) plan was based on the framework for evaluating e-learning proposed by Phillips, McNaught, and Kennedy (2012), which includes the LEPO framework for learning. The focus is to evaluate learning design for 'fit' with achieving the learning objectives and, for the mature GP program design, research the effectiveness and impact of the GP program on students. This paper presents the method and results of a core activity of the EER plan: review data collected during Phases 1 and 2 (pilot and implementation) where the GP program was embedded into two first-year units in the Faculty of Health Science. We describe the method and discuss the results of our analysis in terms of a protocol for a systematic analysis of each data set in terms of ability to inform learning design, the impact and effectiveness of the GP program and usefulness of data in terms of contributing to the development of a diagnostic tool to measure cultural competence.

Keywords: Educational evaluation research; cultural competence; global perspectives; graduate attribute

Introduction

The Global Perspectives (GP) Program is a curriculum initiative at the University of Tasmania (UTAS), designed to teach and assess the global perspectives graduate attribute and equip students for cultural competence and lifelong learning in a global society. The GP Program design was informed by the Learning Environment, Learning Processes and Learning Outcomes (LEPO) framework, presented in Phillips et al. (2012). The LEPO framework is a synthesis of a range of higher education and e-learning research: a "generalized and integrated conceptual framework for learning [that is] pedagogically inclusive" (Phillips et al., 2012, p. 42). LEPO provides a conceptual foundation for rigorous educational research and evaluation of a learning design. The GP Program incorporated each element in the LEPO framework in its design: learning environment, learning processes and learning outcomes, particularly attending to the interrelationships between the elements and the roles of teacher and learners. Using the framework allowed educational designers to choose specific strategies and understand their impacts on different aspects of the learning context. The framework was also the foundation for an evaluation-research plan that integrated evaluation of the design with research into impact and effectiveness.

Kelder, Sondermeyer, Phillips and Rothwell (2012) reported on the rationale and design of a systematic plan for educational evaluation and research (EER) that aligned with the design, delivery and project management of the GP Program curriculum initiative. The realities of a large, multi-disciplinary team busy with design and implementation of the GP program suggested a flexible, opportunistic approach to data collection during the initial phases of the project. This approach was predicated on a commitment to evaluate the data for relevance and suitability for analysis to inform learning design and measure research impact of the GP Program on students.

This paper presents the method and results of a core activity of the EER plan: a review of data collected during Phases 1 and 2 (pilot and implementation) in which the GP program was embedded into two first year units in the Faculty of Health Science (FHS). We describe the method and present the results of our analysis, and discuss the implications for the GP program design in Phase 3 (roll out to first year units for all Schools in the FHS) as well as future data collection and analysis.

The Global Perspectives (GP) Program

The GP Program was designed to support first year FHS students from UTAS to learn and apply knowledge and skills to demonstrate the 'global perspective' or 'cultural competence' graduate attribute. Cultural competence is nominated as a critical graduate attribute for FHS graduates training for clinical professions. The GP Program built on prior curriculum initiatives, directed at students unfamiliar with Australian culture, to prepare them for clinical professional experience placements and interactions with clients. Critically, the vision for the GP Program included a conceptual shift from a 'deficit model' (Sondermeyer, van den Berg, & Brown, 2005)

to a learning design that is embedded as a compulsory, assessable component of curricula for all students.

The intended impact on students who participate in the GP Program is, "cultural competence for life-long learning and work in a global society." The definition of cultural competence was formulated from four major components identified by a literature review of 'cultural competence' and equivalent concepts (Deardorff, 2009).

The LEPO framework conceptualizes learning along three interrelated dimensions: learning environment, learning processes, learning outcomes in the context of learners and teachers interactions (see Phillips et al., 2012, Figure 3.1, p. 27). The opportunities for design of the learning environment were initially constrained by the fact that the GP Program had to be delivered within an existing first year unit, following the traditional lecture/tutorial format. To enhance learner-teacher interaction, the lecture was rebadged 'plenary lecture' and delivered in a highly interactive style. Likewise, the tutorial was named 'tutorial discussion' and included activities designed to prompt discussion, personal reflections and transformational conversations. Given the explicitly challenging nature of the content and potential for negative interactions, the learning design implications included training tutors to manage situations in which students failed to interact appropriately, for example making racist remarks.

Table 1: Constructive alignment (Biggs & Tang, 2007) embedded in the learning design

Cultural Competence Definition	Learning Objectives
An ability to communicate/interact effectively and appropriately with people of different cultures, comprising four components:	Four learning objectives derived from the definition. Each learning objective the focus of a week in the four-part GP program.
a) AWARENESS of one's own cultural worldview;	Student identifies awareness of his/her own worldview in the context of other worldviews;
b) ATTITUDE towards cultural differences;	Student demonstrates a respectful attitude towards cultural differences;
c) KNOWLEDGE and ACCEPTANCE of different cultural practices and worldviews;	Student demonstrates recognition and understanding of different cultural practices;
d) SKILLS (including COMMUNICATION).	Student uses communication effectively and appropriately to enhance intercultural understanding.

In recognition of the interrelationships between learning processes and learning outcomes, a core principle of the learning design to ensure constructive alignment of definition, learning objectives and assessment (Biggs & Tang, 2007) (see Table 1 above). Learning objectives were embedded in the core components of the GP Program and the acted as criteria for decisions on content, structure, learning processes (including sequencing) and supporting resources.

Table 2 sets out the components of the GP Program as it was developed and delivered 2011-2013.

Table 2: Global Perspectives program adapted from (Kelder et al., 2012)

Components	Purpose	Delivery mode	Delivery sequence
Quiz Not assessed	(self)-'consciousness raising' Conversation scaffold for discussion in Module.	Online Via link on LMS Voluntary and anonymous	Administered prior to four-part Module Students notified via LMS email
Module Four instances of **plenary session** followed by **tutorial discussion** Assessed – Exam	Module – each week focuses on one of the four learning objectives (LOs) content related to an LO for that part of the Module discussion and exercises to consolidate and apply the LO to personal and professional practice	Face to Face whole of student cohort interactive plenary session small group tutorial discussion (25 max) Online plenary session recorded and uploaded to LMS.	Plenary session: one hour per week for four weeks Tutorial discussion: one hour per week for four weeks
Workshop Compulsory attendance, not assessed	Preparation for students' Professional Experience Placement (PEP)	Face to Face 30 minutes GP program content embedded in 2 hour PEP workshop	Delivered in following semester

The learning design is structured around three elements: an online *Quiz*, a face-to-face teaching *Module* and, for students who are required to participate in clinical Professional Experience Placement, a *Workshop* component (Kelder et al., 2012). This design explicitly enhanced student engagement with the content and interactions with each other and with the team of teachers. The Quiz, for example, was administered prior to the first plenary lecture as a 'consciousness raising' exercise; the class responses reported in the plenary lecture and subsequently used to facilitate discussion on cultural worldviews.

An Educational Evaluation and Research (EER) plan for the GP Program

The project to develop the GP Program curriculum initiative addressed the question, 'How to design, deliver and manage a curriculum initiative to achieve the 'global perspectives' or 'cultural competence' graduate attribute in health science students?' The current context in higher education is outcomes-focused quality assurance mandated by the Tertiary Education Qualifications Authority (TEQSA) and professional accreditation bodies. Concurrently, a range of university-wide projects are being led by the Student Evaluation, Review and Reporting Unit (SERRU) and the Tasmanian Institute of Learning and Teaching (TILT), for example redesigning student evaluation of units of study (eVALUate) and Course Review Guidelines that embed evaluation into the process. The drive for evidence-based quality assurance at all levels of higher education activity prompted the project team to integrate a parallel stream of activity into the GP Program development that addressed the problem, 'How to design a systematic plan for educational evaluation and research (EER) that is aligned with the design, delivery and management of a curriculum initiative?'

Learning designs, such as the GP Program, have a life cycle over four broad phases of analysis, design, development and implementation. Ensuring rigor and relevance for evaluation in educational contexts is difficult due to the situated and highly contextual nature of educational design and delivery. A systematic and planned approach is required to manage a complex mixture of evaluation and research over a life cycle of evaluative activity: an initial focus on evaluation of learning design in early phases to quality assure the design is necessary before a design is mature and the focus can shift to researching the effectiveness and impact on students (Phillips et al., 2012).

Kelder et al. (2012) provide a high level view of the EER plan over each stage of the design life cycle of the GP Program and for the evaluation of the project management. The EER plan for the GP Program is based on an evaluation-research framework which distinguishes four interrelated, and potentially concurrent, evaluation-research activities: *baseline analysis, design evaluation, formative evaluation* and *effectiveness research* with *project management evaluation* as a separate, related, activity (Phillips et al., 2012). The intended outcomes of the EER plan are: 1) enabling evidence-based improvements to the GP Program design and; 2) providing a reliable and valid evidence base for demonstrating learning outcomes and usefulness of resources. The

tight integration of educational evaluation and research informing the ongoing design and re-design of a learning program is intended to enable longer-term impact research to establish the extent to which students are equipped for future roles in a global society. Formative and summative evaluation of the project management is conducted throughout.

A mixed methods, or hybrid, approach is built into the GP Program EER framework. Ruhe and Boudreau (2011) note, "'…appropriate' assessment of curricular innovations is an *argument* that includes diverse kinds of scientific evidence, as well as the theoretical rationale and the social significance of the innovation" (p. 188). Benefits of designed evaluation include opportunities to improve program design; communication between project team members; identifying additional/alternative types of data to collect, reviewing evaluation procedures, and preventing misunderstandings including with those with oversight of the project (Sanders & Nafziger, 2011[1976]).

The high level plan for EER (Kelder et al., 2012) incorporated the following core features from Phillips, et al. (2012): the LEPO framework, the concept of the 'e-learning design life cycle' with five sequential forms of evaluation-research (baseline analysis, design evaluation, formative evaluation, effectiveness research and project management evaluation). The distinction between 'design evaluation' and 'effectiveness research' is important in terms of timing. The EER plan adopted the principle that research to measure the impact of the GP Program (on students' learning processes and learning outcomes) should not occur until the GP Program had been evaluated from a learning design perspective, refined and established as a mature learning environment for students. The evaluation-research matrix tool supports the selection of research questions appropriate to each stage of the GP Program's design life cycle. Table 3, adapted from Kelder et al. (2012) maps the project phases for the development of the GP program to the stages of a learning design life cycle, with corresponding evaluation-research activities and research questions (Phillips et al., 2012).

Table 3: GP program and EER plan

GP program	**Learning design activity**	**Evaluation – Research activity**	**Research Questions - focus**
Project Phase 0 Life cycle stage 0	Problem description	Baseline analysis	What is the problem and how can we solve it?
Phase 0 Life cycle 1	Design GP program	Design evaluation	How good is the design?

GP program	Learning design activity	Evaluation – Research activity	Research Questions - focus
Phase 1 Life cycle 2	Pilot (Nursing, year one students)	Formative evaluation of the e-learning environment	How can the e-learning environment be improved?
Phase 2 Life cycle 3	Implementation (Pharmacy, year one students)	Formative evaluation of the e-learning environment and processes	How well does the e-learning environment work to support cultural shift?
Phase 3 Life cycle 4	Faculty Health Science roll-out (all year one students)	Effectiveness research into learning processes and outcomes. Impact research on students' thinking and personal / professional behaviours	How effective are the learning processes in generating 'cultural shift' in students? What is the impact on students' capabilities in managing cultural diversity?

Data collected to date – evaluation for usefulness

Phases 1 and 2 of the GP Program included a series of exploratory data collection activities with a view to analysing the data sets for usefulness in evaluation and development of an educational program. For each data set, this paper addresses the question, *Will this data support an answer to the research questions: 1) with a design evaluation focus and; 2) with a focus on impact research?*

The exploratory nature of the data collection included identifying what types of data could be collected; what participant responses were likely and what problems with a data set might emerge in terms of its usefulness for effectiveness and impact research. Another consideration was the workability of collecting the data set. This is particularly important as, practically, data management of large data sets and a large number of data set types, including analysing data, require significant resources. The next section sets out the method for analysing and evaluating the data sets collected during phases 1 and 2 of the GP Program.

Method

The foci of the evaluation of the GP Program were to construct an evidence base to inform the development, and demonstrate the impact and effectiveness, of the learning design. This provided a framework for an opportunistic approach to collect a wide range of data types. Decisions with regard to the category of data gathered were ongoing and emerged as a result of experiences and discussions between project team members. In this exploratory stage several data sets were interrogated to determine their validity in terms of appropriateness for each element of the program evaluation. Validity was determined using a qualitative approach, drawing on the project team's experiences and perspectives.

We deconstructed data drawn from a variety of sources including, data drawn from quizzes developed as a cultural awareness raising exercise and presented pre to the GP Program presentation in two cohorts of first year students (Nursing and Pharmacy). The quiz was not presented to Pharmacy students post completion of the GP program as the Nursing cohort had minimal response rates post completion. Data were gathered from the "My Perspectives Questionnaire" also presented pre and post GP Program presentation to Pharmacy students. This questionnaire was originally designed as diagnostic tool to measure cultural competence levels in undergraduate students. Completion of the quiz and the "My Perspectives Questionnaire" was voluntary. Responses to examination questions completed by Nursing students were examined. These students were given an 'opt out' option in terms of allowing verbatim responses contained in their examinations to be used for research purposes. An essay competition was offered to a cohort of first year medical students *not* exposed to the GP Program. This competition asked the students to write approximately 1000 words focusing on their opinions and experiences of culture. Each entrant in the competition received a small participation gift and a certificate of participation that could be added to their student portfolio. The three best essays were awarded $100 voucher and a certificate of achievement.

We developed a protocol for a systematic analysis of each data set. Elements were analysed by data type (qualitative, quantitative) and in terms of their contextual validity. Data were also considered in terms of ability to inform learning design, and ability to inform the impact and effectiveness of the GP Program. A further focus was to determine usefulness of data in terms of contributing to the development of a diagnostic tool to measure cultural competence. An awareness of cultural biases inherent in any circumstance guided a self-reflective, critical approach by the project team to the evaluation and determination of 'quality' and 'fitness for purpose' of each data set. Students provided informed consent and data were gathered under the approval of the University of Tasmania's Social Sciences Human Research Ethics Committee (*HREC H0012566*).

Results and Discussion

Data were gathered across three first year cohorts within the Faculty of Health Science (Nursing, Pharmacy and Medicine). All student-related information was completed and gathered on a voluntary basis with completion rates ranging from less than 5 percent to 100 percent and varying across each cohort (see Table 4). Four types of data have, to date, been collected with the aim of evaluating the learning design and impact of the GP Program.

Quiz

Year 1 Nursing and Pharmacy students completed a 22-question quiz prior to exposure to the GP Program. This quiz aimed to raise awareness of differing cultural practices and situations by presenting a series of situations/scenarios along with four or five possible responses. The quiz was developed within a forced choice framework where students could not complete the quiz unless they opted for a response presented. If students felt the response options were outside that which they would normally respond they were instructed to respond with the option that was "closest to the response" they would actually have. Demographic data were also collected.

As a 'consciousness raising' exercise that facilitated discussion, the quiz worked very well. However, the use of this data in terms of the measuring the impact of the GP Program is limited. The forced-choice format restricts the validity of responses by assuming that students' responses would fall into the provided categories. Given that all behaviours are cultural artefacts (Krentzman & Townsend, 2008) then decisions made by the project team members in the development of the quiz scenarios and the corresponding possible responses are very likely influenced by their cultural knowledge and experiences. Developing scenario based questions that are measured on linear Likert scales allows for more measured responses. For example, gauging the level of agreement with a statement measured on a continuum ranging from strongly agree to strongly disagree (Kaplan & Saccuzo, 2009). Alternatively, and as a minimum, a further open or "other" option should be made available to respondents. This approach will be considered in future iterations of the quiz.

The intention of gathering pre and post GP Program data from the quiz, while sound in purpose, was not successful. Very few students completed the quiz following the completion of the GP Program. Thus if this exercise and its associated data are deemed necessary in the future then an incentive or inclusion of the follow-up quiz in the assessment requirements of the unit is necessary. Alternate forms of the quiz should also be considered to overcome any learning effects that may carry over from the first completion.

The data gleaned from this aspect of the GP Program has the potential to inform the development of a diagnostic tool to measure cultural shift. Examining the responses to the quiz across the cohorts provided a range of insights. Interestingly, responses to the quiz questions were quite similar across the Nursing and Pharmacy cohorts. At one level this might indicate that the questions are culturally laden with little

difference in responses. However, this might simply be explained by a bias toward socially desirable responses, with considerably fewer responses evident to negatively phrased response choices. One quiz item provides an example of this:

> *A man in the bar of a small hotel has been drinking a lot of beer. Suddenly he collapses and vomits on the floor. If you were nearby, you would:*
> *a. move away;*
> *b. go and get one of the bar attendants;*
> *c. ask him if he'll allow you to help him;*
> *d. call an ambulance.*

Responses to this scenario were overwhelmingly positive with just 7% of Pharmacy students and 3% of Nursing students indicating that they would avoid a person who is intoxicated. The remaining responses indicated a positive helping attitude in this situation. Whether this attitude is inherent in these cohorts or merely a response that is considered appropriate is unclear and highlights the need to consider social desirability when attempting to measure cultural competence. Such data provide critical information for the development of a diagnostic tool that will be able to broadly measure shifts in levels of cultural competence over time.

My Perspectives Questionnaire

The "My Perspectives Questionnaire" was completed by Pharmacy students, in Phase 2 delivery of the GP Program. It consisted of three questions, measured on a 3-point Likert-type scale, focusing on curiosity about other cultures; preparedness for working with people from other cultures and confidence in communicating with people who speak other languages. A fourth open-ended question was also included for students to indicate their understanding of "world view". The majority of Pharmacy students (95%) completed the "My Perspectives Questionnaire" pre the GP program, however this completion rate dropped to 84% post GP program. The resultant data set consisted of 42 valid data points.

Table 5: Means and Standard Deviations of "My Perspectives Questionnaire" data

Question	Mean (out of a possible score of 3)	Standard Deviation
Question 1- Pre	2.27	0.47
Question 1 - Post	2.60	0.50
Question 2- Pre	1.86	0.65
Question 2 - Post	2.33	0.53
Question 3- Pre	1.71	0.67
Question 3 – Post	2.05	0.66

Paired-sample *t*-tests revealed that students indicated that, post the GP Program, they were significantly more curious about other cultures ($t(41) = -3.71$, $p = .001$), significantly more well-prepared for meeting, working with, or caring for people whose culture makes them feel uncomfortable, ($t(41) = -4.60$, $p < .001$), and significantly more confident about communicating with people who speak an language other than their own, ($t(41) = -3.15$, $p = .003$), compared to pre the GP Program. For means and standard deviations see Table 5.

These results seem to indicate an overwhelming success in terms of the impact of the GP Program. Increased curiosity, confidence and preparedness are the outcomes for which the GP Program was designed. Perusal of the comments given to the open-ended question regarding "world view" revealed further positive outcomes. For example, one student responded "*I don't know what that means*" pre GP Program and post responded "*Being open to other peoples' values, cultures and way of life.*" Again, this response clearly shows a shift in cultural awareness, matching with the aims of the GP Program and suggesting that the "My Perspectives Questionnaire" might be useful as a diagnostic tool to measure of cultural competence levels in undergraduate students. However, deconstructing the "My Perspectives Questionnaire" raised some questions in terms of its validity for this purpose.

The questionnaire was not psychometrically validated and, with just three Likert-type items measured on a 3-point scale, it was unlikely to be able to discriminate multiple levels of cultural competence (Kaplan & Saccuzo, 2009). The same small scale questionnaire, presented pre and post GP Program, would also be likely to result in learning effects and thus alternate forms of the questionnaire would have yielded more reliable data. The way in which questionnaire items were presented with answer options referring to curiosity, preparedness and confidence may have positively skewed responses since these traits are considered desirable in terms of cultural competence. Social desirability bias is the tendency of respondents to answer questions in a manner that will be viewed favourably by others and this bias poses problems with all self-report and questionnaire data (Crowne & Marlowe, 1960; Janda, 1998). In a recent and comprehensive review, Krentzman and Townsend (2008) provided a summary of existing scales and suggested processes for the development of cultural competence scales. They noted the need to consider socially desirable responses and suggest that a valid social desirability measure is completed alongside any cultural competence scale. Krentzman and Townsend (2008) point out the need to consider that all scales are, in fact, cultural artefacts and thus there is a need for a multidisciplinary, multicultural approach to their development. Thus the "My Perspectives Questionnaire" is not valid or useful in terms of measuring impact of the GP Program and further consideration will be given to developing a psychometrically validated diagnostic tool that will accurately measure changes in a range of cultural competencies including knowledge, skills, awareness, attitudes and behaviour.

Assessment

Assessment is a critical and high value data set that informs learning design and is a measure of impact and effectiveness of the approach taken to implement curricula (Phillips et al., 2012). Phase 1 and 2 of GP program focused on raising awareness and providing information associated with knowledge of different cultural practices and world views. This learning was assessed by inclusion of three questions in the examination that represented 10% of the overall examination weighting. The assessment questions were aligned to introductory knowledge level learning objectives (Anderson & Krathwohl, 2001; Bloom, Engelhart, Furst, Hill, & Krathwohl, 2001). Assessment of the GP Program was limited to one cohort (Nursing) due to constraints on the Pharmacy examination design, thus comparison of cohort assessment outcomes was not possible.

The assessment responses were analysed for degree of achieving learning outcomes within a framework of Bloom's (1956) taxonomy. The assessment questions were pitched at the lower levels of 'Understanding' and 'Knowledge' and with an overall 62% mark on these questions providing an indication that on average students have grasped the required knowledge at an introductory level. This confirms that the learning design is appropriate at this level and provides a foundation from which further, higher level, learning objectives and relevant exercises and assessments can be designed. Future development of the GP Program will focus on progressing learning within the cognitive domain (teaching and assessing students' application, analysis and synthesis of cultural information and understanding) and also developing their affective (attitudes) and psychomotor (behaviours) skills (Anderson & Krathwohl, 2001; Bloom et al., 1956). When designing higher levels of learning and associated assessment, consideration could be given to more complex thinking; evidence applied in practice (e.g., clinical students) and supervisor assessments.

Students were given the option to allow their examination responses to be used for research purposes and, at the time of writing, 24 have given consent to have their data included the evaluation process. This has allowed a more in depth scrutiny of responses and this snapshot of data has provided further insight that may inform the learning design of future elements of the GP Program. Although, on average, students provided responses adequate to pass this element of assessment, some failed to answer one or more questions. This may suggest a lack of understanding or perhaps a pragmatic response by students to nominally weighted examination questions. Interestingly, some responses were very similar, perhaps indicating a level of rote memory strategies supporting knowledge in this area. If this is the case then a revised learning design should focus learning objectives and aligned assessments in a more applied manner.

Comparison of student outcomes across cohorts for consistency is an important consideration in all learning. It is particularly important in the instance of the GP Program, since this is a faculty wide initiative and will be rolled out to all Schools

across multiple campuses and many staff will be involved in teaching and assessing various aspects of the course. The inability to accomplish standard assessment in Phase 2 of the GP Program roll-out has highlighted the need for a mandatory assessment component to be included in each unit of study that embeds the four-week GP Program.

Essay

Given the limited usefulness of the data and the recognition of the influence of social desirability biases inherent in the "My Perspectives Questionnaire" project team members decided to trial an essay competition with first-year medical students. Roll out of the GP Program to medical students will occur in Phase 3 thus the aim of the competition was to gain an insight into students' perspectives on culture prior to exposure to the GP Program. This exercise was implemented in an effort to optimise honest, open answers and minimise elements of social desirability in responses. The invitation to students directed them to reflect on personal understandings of culture, discuss a range of areas and provide evidence to support observations. Students were asked to consider their concept of culture; their knowledge about people unlike themselves; their thoughts on how their behaviour impacts others; their place in the world and; how they talk about, or talk to, people who are different from themselves.

The invitation was offered to 115 students and, despite substantial incentive, only five students responded. Evaluation of the essays was undertaken by two raters in two parts: independent marking guided by a rubric to identify evidence of awareness, knowledge, skills, attitudes and behaviours associated with cultural competence, followed by discussion of rater's individual interpretations of essay responses. Initially, given the low response rate, this essay competition was considered unsuccessful. However, closer examination of the essays revealed a very distinct and rich source of information. In just five responses there was clear evidence of high levels of cultural acceptance, awareness and experience. Likewise evidence of a range of interpretations of culture and how it impacts students' lives in educational and personal environments was presented. One essay, for example, described the way in which understandings of privilege impacted and were dealt with by some students within the cohort. In discussion, the raters understood this as a clear example of cultural understanding that extends beyond mainstream concepts of culture and a reminder of the nuanced nature of the subject matter of the GP Program. This productive understanding, that every aspect of the GP Program is in fact a cultural artefact (Krentzman & Townsend, 2007), will inform and support the development of a diagnostic tool that is as broadly applicable as possible.

Given the sensitive nature of some perceptions and understandings of culture, the way in which these data were gathered provided a forum for students to provide their opinion without the risk of impacting their marks. Voluntary entry to a competition outside assessment requirements, while resulting in limited responses, provided a freer environment for responses. This possible restriction associated with

assessment must be taken into consideration when analysing qualitative responses of, what are traditionally, high achieving students.

A significant and important outcome of the evaluation of this data set was that, despite the use of a standardised rubric, raters had very different responses to three of the essays. This, along with unanticipated responses, prompted reflection on raters' assumptions, ideas, and the validity of the rubric. This process also identified aspects of the GP Program that maybe cultural artefacts of the composition of the project team who decided the content of the program. Future development of the core curriculum will be informed by the deconstruction of the essay insights. Table 6 provides a summary of the data sets analysed according to the protocol.

Table 6: Summary of evaluation of data sets

Data	QL	QT	Contextual validity	Informs learning design	Informs impact and effective-ness	Diagnostic tool
Quiz	? ×	✓ ×	Raising awareness; Informing item development of diagnostic tool	Embedded in the learning design	Supported development of diagnostic tool	Student/staff submission of items for question bank; Some items to transfer from Quiz to tool
My Perspect-ives Q'naire	✓	✓	? High social desirability issues ✓Pre/post	Not useful	Highlights need for validated diagnostic tool	Must address social desirability
Assess-ment	✓	✓	Aligned to learning objectives; Exam; portfolio options	Blooms Taxonomy (BT) analysis informs design for additional learning	Can measure using BT as criteria (limited; dependent on assessment type)	Deconstructed assessment will have elements that inform item development

Data	QL	QT	Contextual validity	Informs learning design	Informs impact and effective-ness	Diagnostic tool
Essay	✓	×	Low social desirability influence; Low response rate.	Articulated questions and understand ings of culture that not yet addressed by GP program;	Conclusion 'not useful' (low response rate). Decon-structed data set showed high quality info that challenged project team assumptions	Deconstructed text will have elements that inform item development

Conclusion

To date, the EER plan has guided the formative evaluation of the GP program design and implementation. The data collection process was exploratory and opportunistic in Phases 1 and 2 due to the situated and contextual nature of the design process for the GP Program. We have developed a protocol to interrogate and evaluate the data sets from different cohorts and for different aspects of the GP Program (design and delivery). This was a necessary and useful mechanism to evaluate the data for quality and fitness for purpose. The outcomes of this evaluative activity are recursive, in that they also provide a lens for evaluating the intended purposes of the data collection. For example, the research questions have evolved in response to unexpected insights from this meta examination of the data sets.

Results indicate that assessment items are a critical data set for determining achievement of learning objectives. However, assessment must be carefully designed to enable pedagogical measurement of student learning. In particular, assessment tasks must test students' cognitive (knowledge and understanding), affective (attitudinal) and psychomotor (behaviour) (Anderson & Krathwohl, 2001; Bloom et al., 1956) gains in response to the GP Program. This suggests the need for a framework for assessment design to ensure breadth and depth of learning for each cohort and year level exposed to the GP program.

The EER framework includes a long-term strategic aim to measure effectiveness and impact on students. The evaluation of the data sets to date confirmed that the development of a rigorous and validated diagnostic tool will be problematic for a number of reasons, conceptual and methodological. However, the cumulative data set indicates that a rigorous, multi-methods approach and multiple sources of data will provide a sound foundation from which a psychometrically validated diagnostic tool can be developed.

The evaluative activity also confirmed that the GP Program and its various data sets are themselves a cultural artefact. Each element reflects the political drivers, responses and resistances, values, experiences, framing of experiences and knowledge of the project team members. Thus, the "cultural competence" definition adopted by the project has served its purpose well. It has been the guiding principle for identifying learning objectives and designing assessment. It has also provided a lens for each member of the project team, in whatever role, to reflect on and analyse his or her contributions to the content and the evaluation design (what questions we ask) and to ensure that each component models the definition for cultural competence developed by the educational development team.

References

Anderson, L. & Krathwohl, D. A. (2001). *Taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of Educational Objectives.* New York: Longman.

Biggs, J. & Tang, C. (2007). T*eaching for quality learning at university. Third edition*. Berkshire, England: Open University Press.

Bloom, B.S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals; Handbook I: Cognitive Domain* New York: Longman.

Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349-354.

Deardorff, D. (2009). *The Sage handbook of intercultural competence.* Thousand Oaks CA: Sage Publications

Janda, L. H. (1998). *Psychological testing: Theory and applications.* Massachusetts: Allyn & Bacon.

Kaplan, R. M. & Saccuzo, D. P. (2009). *Psychological testing: Principles, applications, and issues. Seventh edition.* Belmont CA: Wadsworth, Cengage Learning.

Kelder, J-A., Sondermeyer, J., Phillips, R., & Rothwell, A. (2012). *Designing evaluation and research into educational initiatives: The Global Perspectives Program. Ascilite 2012*. Wellington, NZ.

Krentzman, A. R. & Townsend, A. L. (2008). Review of multidisciplinary measures of cultural competence for use in social work education. *Journal of Social Work Education, 44,* 7-31.

Phillips, R., McNaught, G., & Kennedy, C. (2012). *Evaluating e-learning: guiding research and practice*. New York: Routledge.

Ruhe, V. & Boudreau, J. D. (2011). The 2011 program evaluation standards: A framework for quality in medical education programme evaluations. *Journal of Evaluation in Clinical Practice, 23,* 187-200.

Sanders, J. R. & Nafziger, D. N. (2011[1976]). A basis for determining the adequacy of evaluation designs. *Journal of Multidisciplinary Evaluation, 7,* 44-78.

Sondermeyer, J., van den Berg, K., & Brown, J. (2005). Making nursing education culturally competent: moving beyond the deficit model. *International Conference on Critical Discourse Analysis: Theory into Research*, 669-676.

Student Evaluation, Review and Reporting Unit. Website available: http://www.utas.edu.au/student-evaluation-review-and-reporting-unit/our-serru-team Accessed 16th November 2012.

Tertiary Education Quality and Standards Agency. Website available: http://www.teqsa.gov.au Accessed 16 November 2012.