ASCILITE SHAPING THE FUTURE OF TERTIARY EDUCATION
Show Me The Learning
Adelaide, Nov 27-30, 2016

Evaluation of a learning outcomes taxonomy to support autonomous classification of instructional activities

Mark McMahon
School of Arts & Humanities
Edith Cowan University

Michael Garrett
CingleVue International Pty Ltd

With an increased focus on assuring the quality of student achievement in higher education, there is a commensurate need for tools to assist academics in understanding the nature of assessment and how it can provide evidence of student learning outcomes. This paper describes research conducted the Instructional Activity Matrix; a taxonomy that was developed as the basis of a learning support tool, *Maestro*, that automatically analyses outcomes and assessment statements to show the cognitive level and nature of knowledge inherent in them. Findings indicate that the matrix is a valid tool for defining the nature of learning outcomes and had value in clarifying the nature of assessment and outcomes. However, issues identified with the inherent ambiguity of some instructional statements and their contextually-laden language provided insights into how Maestro will need to be refined to provide appropriate support for teachers, with a range of experience across multiple disciplines.

Keywords: learning outcomes, taxonomies, learning management systems, assessment

Introduction

The Australian tertiary education sector is currently undergoing a period of transformation, characterised by an increased regulatory impost on universities with regard to the demonstration of quality standards across a range of higher education functions. In 2009, following the Bradley Review of Australian Higher Education (Bradley, Noonan, Nugent, & Scales, 2008) and accommodating the expansion of the sector through the establishment of the Tertiary Education Quality and Standards Agency, TEQSA, the Australian government heralded a 'new era of quality in Australian Tertiary Education' (DEEWR, 2009). This has resulted in the a revised Higher Education Standards Framework, which from January 2017 defines the thresholds that universities and other tertiary institutions need to achieve for provider registration, course accreditation and qualifications (Birmingham, 2015).

This formalization of standards has evolved from the recent extension of national outcomes for university degrees through Australian Qualifications Framework to all levels of post-secondary education, including undergraduate and postgraduate qualifications and higher degrees by research (Australian Qualifications Framework Council, 2013), which has led many universities to develop course or program level outcomes as well as specifications for evidence needed to demonstrate these. However, simply having a set of learning outcomes does not ensure quality, nor does it necessarily lead to improved outcomes for graduates. Thompson-Whiteside (2012) noted the risk of national standards restricting innovation and good practice and argued, 'academic staff need to individually and collectively, within their disciplines, have the autonomy to set and assess their own standards' (p. 35).

While empowering teachers to be the sentinels of quality in higher education is a noble ideal, it does raise questions about the ability, and perhaps more importantly the capacity, of academics in terms of their available time to assure this quality. While there are a number of systems that have been developed to assist in the management of learning outcomes such within the Blackboard Learning Management System (Blackboard Inc., 2016), these predominantly take the form of instruments to map outcomes to assessments and act as repositories for evidence. This compliance focus may make the process of managing outcomes and assessment easier but there is little to assist the educators themselves in understanding the nature of their outcomes and whether their approaches to assessment are achieving those.

This type of support is the focus of *Virtuoso*, an enterprise learning and instructional support platform that is currently under development to support quality learning in secondary and tertiary educational environments. This paper reports on research undertaken to design and develop an autonomous system within *Virtuoso* for classifying instructional activities such as outcomes and assessment according to their cognitive level and type of knowledge. This system, *Maestro*, makes use of an underpinning learning outcome taxonomy, the Instructional Activity Matrix.

The Instructional Activity Matrix and Autonomous Classification

Identifying learning outcomes is an important initial process in developing assessment and approaches to learning. The use of taxonomies allow for classification to help distinguish the nature of these (Jonassen, Tessmer, & Hannum, 1999). One key aspect of taxonomies is that they are hierarchical. Those classifications that sit at the top of the taxonomy are more general, inclusive, or complex, with lower level classification often identifying subordinate or prerequisite attributes (Jonassen & Grabowski, 2012). In the development of the Instructional Activity Matrix, a range of learning taxonomies were evaluated for their ability to embrace a broad range of cognitive learning outcomes. These included Bloom's Taxonomy, Gagne's Taxonomy, Merrill's Performance-Content Matrix, the Structure of Observed Learning Outcome (SOLO) taxonomy, and the Revised Bloom's Taxonomy.

Perhaps the best known taxonomy of learning outcomes is that proposed by Bloom (1956). Its six categories in the cognitive domain progress from Knowledge to Comprehension, Application, Analysis and Synthesis, through to Evaluation. Gagne's Taxonomy of Learning Outcomes (Gagne, 1985) differs in its focus on epistemology rather than cognitive levels, with fixed and inert Verbal Information through to Concrete and Defined Concepts through Rules and Higher Order Rules to Cognitive Strategies, the last of which supports the acquisition of new forms of knowledge independently.

The main difference between Bloom and Gagne are that the former defines cognitive levels at which something can be understood, while the latter classifies what is to be understood in terms of its type of knowledge. More contemporary taxonomies embrace both of these dimensions. Merrill's Performance Content Matrix (Merrill, 1983), classifies outcomes according to student performance as well as subject matter content. Student performance classifications include Remember-Instance, Remember-Generality, Use, and Find, while subject matter content classifications incorporate facts, concepts, procedures, and principles. Learning outcomes can thus be classified using two separate dimensions and in multiple cells of the Performance-Content Matrix. While the Structure of the Observed Learning Outcome (SOLO) taxonomy (Biggs, 1989) maintains this complexity, the classifications themselves are more holistic, emphasizing the observation of student learning cycles to describe the structural complexity of a particular response to a learning situation through five different levels: prestructural, unistructural, multistructural, relational, and extended abstract. This makes it particularly useful for developing holistic assessment rubrics but by aggregating cognitive levels and epistemological types into types of responses it makes it more difficult to classify individual instructional statements.

The Revised Bloom's Taxonomy (Anderson et al., 2001), however, embraces the bi-dimensionality of Merrill, following a Knowledge dimension that provides similar epistemological categories to those suggested by Merrill, with the addition of a Metacognitive Knowledge category. The Cognitive-Process Dimension effectively updates Bloom's original taxonomy, emphasising meaningful active processing of new knowledge where Knowledge becomes 'Remember', Synthesis is 'Create' and so on, with the latter of these now privileged as the most complex form of processing, above 'Evaluate'.

The synthesis of these taxonomies resulted in the Instructional Activity Matrix (McMahon & Garrett, 2015). Like Merrill and Revised Bloom, it embraces both cognitive processing (Remember, Understand etc.) and types of Knowledge (Factual, Conceptual etc.). The Cognitive Processing dimension maintains the contemporary call to action of Revised Bloom, while the Knowledge Dimension integrates Principle Knowledge from Merrill as well as the Metacognitive Knowledge of Revised Bloom. These are distinguished from each other in that a principle, while more complex than procedural knowledge, is nevertheless of lower order than metacognitive knowledge in the latter's capacity to generate principles through higher order skills of abstraction and generalization.

The final matrix is presented in Table 1. It provides 30 possible individual classifications for categorising instructional activities, augmenting the 24 possible classifications of the Revised Bloom's Taxonomy through the addition of the principle knowledge classification within the Knowledge Dimension. Each cell within the Instructional Activity Matrix is the intersection of the Cognitive-Process and Knowledge Dimensions that describes the specific cognitive processes and subject-matter content involved.

Table 1: The Instructional Activity Matrix

	1. Remember	2. Understand	3. Apply	4. Analyse	5. Evaluate	6. Create
a. Factual	1a. Recall association	2a. Specify features	3a. Utilise fact	4a. Determine features	5a. Check factual accuracy	6a. Generate factual representation
b. Conceptual	1b. Recognise concept	2b. Characterise concept	3b. Enlist concept	4b. Examine concept	5b. Consider concept	6b. Evolve concept
c. Procedural	1c. Recall procedure	2c. Clarify procedure	3c. Execute procedure	4c. Scrutinise procedure	5c. Critique procedure	6c. Devise procedure
d. Principle	1d. Recognise principle	2d. Explain principle	3d. Relate principle	4d. Investigate principle	5d. Validate principle	6d. Discover principle
e. Metacognitive	1e. Recognise learning fundamentals	2e. Comprehend learning processes	3e. Implement learning strategy	4e. Explore cognitive processing	5e. Assess learning performance	6e. Develop learning abstraction

The Instructional Activity Matrix enables any instructional statement, whether framed as a learning outcome or an assessment task to be classified according to the appropriate forms of knowledge and cognitive level required to demonstrate achievement. So an outcome such as *Name the parts of the respiratory system* involves remembering factual information (1a) while *Design an infographic that shows how the respiratory system works* would require users to evolve a concept, thus intersecting 6 (Create) and b (Conceptual Knowledge).

As with any taxonomy, the Instructional Activity Matrix operates according to a number of assumptions. It assumes that one can characterise human knowledge and activity as discrete cognitive states (Jonassen et al., 1999) and that such states can be identified, specified, and measured reliably and validly (Jonassen & Grabowski, 2012). These principles underpin the implementation of the automated system, where to allow for the classification of instructional statements, each cell in the Instructional Activity Matrix is matched with classification terms within a curated lexicon. The lexicon links verbs and nouns with each of the 30 classification cells of the Instructional Activity Matrix, with verbs mapped to columns in the Cognitive-Process Dimension and nouns mapped to rows in the Knowledge Dimension. Analysing each word in a learning statement using Part-of-Speech (POS) tagging, the automated system thus looks for matching verbs within the lexicon and if it finds them notes the corresponding cells within a given column. This process is then repeated for nouns within the lexicon to identify matching cells within a given row. A final classification for the learning artefact is derived from the intersection of matching cells. Complex instructional tasks, such as those comprised of lower order sub-tasks, can be accommodated via classification in more than one cell.

The Instructional Activity Matrix has been embedded in the *Virtuoso* enterprise learning and instructional support platform currently under development. It operates as an autonomous classification tool, *Maestro*. *Maestro* is able to parse statements that define learning activities, assessments, and learning outcomes, where there is some reference to a task that students are required or should be able to undertake within certain parameters or within a specified context. It then processes those statements using a curated lexicon of verbs and nouns that are aligned to one or more cognitive levels and types of knowledge.

Figure 1 shows how *Maestro* has been embedded into a Beta build of *Virtuoso*. The menu lists outcomes defined by the Australian Curriculum and the example on the main screen demonstrates how *Maestro* has automatically classified one outcome from the Year 9 History Curriculum, displaying the outcome as a heatmap, demonstrating the extent to which the statement meets the various classification cells of the Instructional Activity Matrix. The highest number shows that the statement has the strongest correlation with the intersection between *understand* and *principle* indicating that the outcome has a focus on explaining principles.

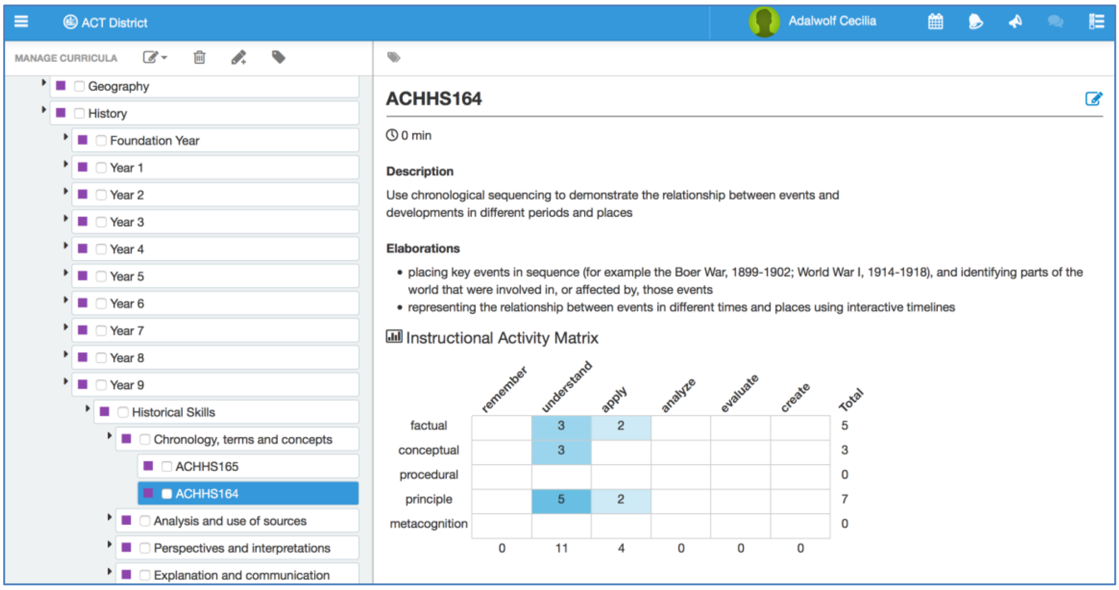

Figure 1: The Instructional Activity Matrix embedded in a beta version of *Virtuoso*

Previous research (McMahon & Garrett, 2016) has compared heatmaps of classifications undertaken by *Maestro* and human educators. Findings have indicated that there is some correspondence in terms of how humans classify compared to the autonomous classifier but also some tangible differences. To further interrogate how individuals used the Instructional Activity Matrix to classify outcomes in comparison to *Maestro*, research was undertaken to understand teachers' experiences with the taxonomy through a structured classification activity. This paper explores how teachers used the Instructional Activity Matrix and its value in terms of:

- Existing skills of teachers to understand the nature of their learning outcomes
- Ability to classify a range of outcomes with the Instructional Activity Matrix across different disciplines

Method

In order to explore teacher application of the Instructional Activity Matrix, a group of 10 teachers were recruited from amongst six schools located in New South Wales. Participants teaching experience ranged between 2 and 30 years in both primary and secondary school capacities across a range of learning areas. The ages of participants in the sample spanned 25 to 54 years old, with 9 females and 1 male.

Each participant in the study was tasked with manually classifying a series of learning outcomes taken directly from the Australian Curriculum using one or more cells in the Instructional Activity Matrix. To prepare participants for this task, the researchers led a workshop which provided background information on taxonomies of learning outcomes and the Instructional Activity Matrix and demonstrated the classification process in detail using numerous worked examples. Each worked example was explored in detail through group discussion where participants compared different perspectives and classification outcomes to clarify their understanding of the classification process. This also included the comparison of manual classifications to autonomous classifications provided by the *Maestro* system. Participants were given the opportunity to classify learning outcomes that they selected prior to attending the workshop with the researchers on hand to provide feedback and assistance to each participant as required.

After completing the workshop, participants were required to classify a series of 129 learning outcomes using the Instructional Activity Matrix. For each learning outcome, participants identified a classification using one or more cells within the matrix to denote the cognitive processes and types of knowledge involved. The first 29 learning outcomes were selected by the researchers from the Foundation to Year 10 section of Australian Curriculum within the subject areas of History, Science, Mathematics, and English. The remaining 100 learning outcomes were selected by the participants themselves from the New South Wales Curriculum. All learning outcomes were expressed in the form of an instructional statement describing the knowledge, skills, and competencies that students are expected to acquire (e.g. *Compare and order duration of events using everyday language of time)*. Once participants had classified all 129 learning outcomes, they completed an online survey designed by the researchers as a simple objective means to elicit participants' backgrounds and experience in using the Matrix. This was followed by a phone interview with the researchers.

Survey Results

With 10 participants, Likert scale responses did not provide a statistically significant sample. Nevertheless, there were noticeable trends in terms of how participants felt they had the prerequisite skills and knowledge, their experience of the workshop and the value they attributed to and reported skills in classifying outcomes according to the Instructional Activity Matrix. These results are summarised in Table 2, showing the survey questions, the most common response, the average response (from 1 to 5 in terms of level of agreement) and the percentage agreement, discounting neutral responses and comparing the number of 4s and 5s with the number of 1s and 2s.

Table 2: Instructional Activity Matrix Survey Responses

No.	Question	Mode	Average	% Agree
1	I have well developed English language skills.	Strongly Agree	4.7	100
2	I have a good understanding of grammar and sentence structure.	Agree	4.1	100
3	I can effectively identify the nouns, verbs, and adjectives within sentences.	Agree	4.4	100
4	I have an extensive vocabulary and understand the meaning of a wide variety of words.	Agree	4.2	100
5	I was familiar with taxonomies of learning outcomes prior to attending the workshop.	Agree	4	100
6	I had used taxonomies of learning outcomes as part of my teaching prior to attending the workshop.	Agree	4.3	100
7	The workshop improved my understanding of taxonomies of learning outcomes.	Agree	3.8	70
8	I understood the purpose of taxonomies of learning outcomes after completing the workshop.	Agree	4.3	100
9	I understood the purpose of the Instructional Activity Matrix after completing the workshop.	Agree	3.9	70
10	I understood the relationship between cognitive processes and types of knowledge in the Instructional Activity Matrix after completing the workshop.	Agree	4	90
11	I was able to classify learning outcomes using the Instructional Activity Matrix during the workshop.	Agree	3.8	80
12	The group discussion improved my understanding of the Instructional Activity Matrix during the workshop.	Undecided	3.7	50
13	The workshop provided with me with the necessary knowledge and understanding to classify learning outcomes using the Instructional Activity Matrix.	Agree	4.1	90
14	I understood the process for classifying learning outcomes using the Instructional Activity Matrix after completing the workshop.	Agree	4.1	90
15	I felt confident to proceed to Phase 2 of the study after completing the workshop.	Agree	4.1	80
16	I was effectively able to interpret the learning outcomes I needed to classify.	Agree	4.1	90
17	I was effectively able to classify the 30 learning outcomes that were provided to me from the Australian Curriculum.	Agree	4	90
18	I was effectively able to classify the 100 learning outcomes that I selected from the NSW Curriculum.	Agree	4	90
19	The classifications that I obtained using the Instructional Activity Matrix made sense to me.	Agree	4.1	90
20	There was a clear relationship between each learning outcome and the corresponding classification I obtained using the Instructional Activity Matrix.	Undecided	3.4	40
21	The classifications I obtained using the Instructional Activity Matrix accurately reflected the nature of the learning outcomes.	Agree	3.4	50

22	I experienced difficulty classifying the learning outcomes.	Occasionally	3.3	30
23	I encountered learning outcomes that I was unable to classify.	Occasionally	2.2	10
24	I consulted a dictionary to check the meaning of words when classifying learning outcomes.	Occasionally	2.3	10
25	I became more effective at classifying learning outcomes as I progressed through the collection.	Agree	3.9	80
26	The Instructional Activity Matrix is an effective tool for classifying learning outcomes.	Undecided	3.2	40
27	The Instructional Activity Matrix could be used to classify learning outcomes that relate to any school subject.	Undecided	3.2	30
28	I developed a better understanding of each learning outcome after classifying it with the Instructional Activity Matrix.	Agree	3.6	60
29	Learning outcomes that have been classified according to the Instructional Activity Matrix are useful to teachers.	Undecided	3.1	30
30	In the future, I would consider using the Instructional Activity Matrix to inform my teaching.	Undecided	3.2	30

As can be seen, participants reported an overall strong level of pre-existing skills in terms of their English language and understanding of grammar and sentence structure, with 100% agreement to questions 1-4, which related to their language skills. Similarly, participants rated their knowledge of learning outcomes highly; with all saying they were familiar with learning outcomes taxonomies and had used them to some extent in their own teaching.

Despite this level of self-reported familiarity with learning outcomes, there was an overall positive response to the workshop in terms of its ability to add to their understanding, with the majority agreeing that the workshop improved their understanding of taxonomies and that they felt able to classify learning outcomes after having completed it.

Weaker but still positive results were found in participants' perception of the Instructional Activity Matrix itself in terms of its capacity to support effective classification and their abilities to use it. The vast majority of participants felt that they were able to classify outcomes from both the Australian Curriculum and the New South Wales State Curriculum and that their classifications made sense to them. However, there was some uncertainty as to whether the Instructional Activity Matrix accurately reflected the nature of the learning outcomes and whether there was a clear link between the outcomes they were classifying and the classifications themselves. This was evident in the fact that several had difficulties in classifying outcomes and some encountered learning outcomes that they felt unable to classify. Despite their high level of reported skills in English and learning outcomes it is also evident that some required a dictionary at times. Nevertheless, the process seemed to get easier the further into it they went. More participants also found that they developed a better understanding of the learning outcomes by undertaking the activity. This is notable given their perceived comfort with both New South Wales and Australian Curriculum outcomes.

Overall, responses were more muted in terms of the applicability of the Instructional Activity Matrix, with most undecided as to its value and whether they would use it in their teaching. Written responses to the open ended question as to why varied greatly. For some, it was highly valuable – 'it helps to ensure that I am catering for a range of thinking skills, including higher order, and giving my students a range of opportunities to learn and represent their learning'. Others pointed to the contextually-bound nature of learning outcomes and how that can create ambiguity – 'I think it will be hard to cover the meaning of all words with the Matrix to different outcomes. I think describe in one subject area might mean something different in another subject area.' One response spoke about the practical aspect of lesson planning, where teachers develop a lesson intention with subordinate success criteria. In that context, curriculum outcomes themselves are marginalized in favour of more granular and practical ways to identify and gather evidence of student achievement.

While these results do not impact necessarily on the validity of the Instructional Activity Matrix, nor its effectiveness as a tool for classifying outcomes, the stated goal of being able to develop assessment where teachers' understandings of student achievement can be calibrated with the feedback provided by *Maestro* means that these issues warrant discussion. The next section draws recommendations for continuing the research by exploring these experiences, by triangulating these results with the deeper responses from individual interviews.

Discussion

In comparing the above results to interview responses, a number of salient themes evolved that can be used to inform future iterations of *Maestro* and how it can be used in a practical education setting. Specifically, these concern the differences between human and autonomous classification, issues around how learning outcomes themselves have been defined as the basis of curricula, and the potential for going beyond the classification of outcomes to identify the ways in which they are evident through the assessment of learning that takes place in the school setting. Each is discussed in turn.

Human vs autonomous classification.

All participants described the process of classifying outcomes as time consuming, to the extent that two specifically described the process of classifying the 29 outcomes from the Australian Curriculum in combination with the 100 that they picked from New South Wales as 'tedious'. Two participants described the classification process as 'hard' though several reinforced the survey finding that it did get easier as they went along. The actual time they took varied from approximately three to five hours, with several of them setting aside a whole day to undertake the activity. The biggest concern was the sheer number of outcomes they were required to work with. While this would not represent an authentic activity for most teachers (who would likely be working purely within their discipline and with a subset required for their specific cohort), it nevertheless identifies workload implications should teachers be required to manually classify both assessment and curricular outcomes with a view to assuring student achievement. In this sense, Maestro is highly efficient automatically processing statements is almost instantaneous and a computer never tires.

Several of the participants ended up undertaking the classification activity with others, including other participants of the research. This was perceived as highly valuable with one participant stating, 'I was able to talk about it while I was doing it. So it was good to get a different perspective as to someone else's thinking'. Another went further, arguing, 'it would have been good maybe to touch base with the other teachers who were working through the same process. That way we could be sure we were doing the right thing, though I did have one of the other participants at [the school].' This social negotiation of understanding was a powerful mediator, particularly, as one participant observed, when cross checked with the handouts and examples provided as part of the original workshop.

This highlights the potential of *Maestro* as a calibration tool. Individual teachers who may not have access to others to compare their interpretations would be able to check their understanding of an instructional activity with the classification performed by *Maestro*. While it may not fully replicate the social construction of knowledge, it would be a highly efficient source of a potentially alternative point of view.

This does raise an issue with regard to the role of *Maestro* as a teacher support. The participant who in interview expressed the least enthusiasm for the project identified specific issues with the ambiguity that was inherent in the workshop and the classification process:

> I would have liked more information and feedback in relation to the accuracy of the classification of the outcomes… more feedback when discrepancy between classifications (automated implementation in *Virtuoso* vs. people). [There was] no real discussion as to situations in which the software and participants didn't have consensus.

This participant described this ambiguity as a weakness of the workshop, requesting 'greater clarification regarding accuracy of classifications, more detailed feedback when there wasn't an obvious consensus'. This need for a single 'correct' answer is counter to the role of *Maestro* as performance support tool rather than a prescriptive one. Nevertheless, it would likely represent the attitude of many teachers and would require significant professional development in the nature of *Maestro* and its role in planning assessment. One inescapable finding was that the process of interpreting learning outcomes had an inherent subjectivity, and this was exacerbated by the ambiguity manifest in some of the outcomes themselves.

Shortcomings of learning outcomes as they pertain to different disciplines

In interview, participants were asked whether there were any outcomes that they struggled with. Overwhelmingly, the strongest response was for those outcomes from the English curriculum, with all participants mentioning that these were difficult to classify. The reasons related to the length of the statements, the tendency to include multiple outcomes in a single statement, and the ambiguity caused by common terms having specific contextual meanings. One example mentioned was the word 'discuss', which ostensibly describes a communication process but in the English discipline incorporates a range of critical and analytical skills.

Two participants were teachers of English and saw this as an inherent aspect of the discipline, with one arguing that that you could never disaggregate these outcomes as the subordinate cognitive processes are ineluctably linked. Such a statement may reflect the holistic nature of English assessment but it does problematise assessment in terms of identifying distinct criteria and measures of success. Both English and History were described by one participant as ‘abstract’.

The fact that one participant stated that it was ‘hard to work out what they’re looking for’ in the English outcomes is a reminder that they are often written by discipline experts, the result being statements that are bound to their discipline context. While it is obviously appropriate for disciplines to develop their own outcomes, it does question the transferability of language used between domains. The survey results showed that participants considered themselves to be well attuned to the outcomes in their own discipline but in interview all participants agreed that they found their own outcomes easier to classify than those of other subjects. The discipline-specific nature of outcomes also appeared to translate to how individuals engaged in the classification process. Both teachers of English took quite a holistic approach to classification. One described how she embraced the complexity of the outcomes and then tried to narrow it down in terms of what skills are ‘privileged’ in the outcome. The process for her proved to be an interesting critical experience:

> That is what I found valuable because I haven’t really read the outcome with this level of depth or understanding before. It really got me to critique the outcome and I thought, wow these outcomes are really asking me to do so much more than just teach grammar’

The approach of one of the science teachers, however, was much more methodical. When asked if she felt the process was subjective, she responded, ‘if you are of a mathematical or scientific disposition, probably not’. This is reflected in her own classifications where she would follow a similar approach to *Maestro*, in which individual verbs and nouns had discrete definitions and when she came across them repeatedly she would classify them in a consistent manner regardless of the context of the outcome.

Since *Maestro* is contextually neutral in its autonomous classification, this highlights both an advantage and a disadvantage of the system. By following a simple dictionary, its interpretation of vocabulary is not coloured by discipline bias, however it lacks the capacity to provide a nuanced classification dependent on which discipline in which the outcome sits. One way in which this could be improved would be by integrating discipline specific definitions of terms within *Maestro*. This would allow some generic words to be better honed to a subject so a verb like ‘find’, for example, may reflect how it is used to resolve problems in Mathematics but promote innovation in Design and Technology.

Outcomes vs. assessment

One key finding both in surveys and interviews was that participants found the Instructional Activity Matrix to be a valid instrument for classification. While some mention was made in interview of the difficulty of understanding the nature of procedural knowledge and the breadth of outcomes inherent in conceptual knowledge, one participant noted:

> As I’ve been doing this I’ve realized how these outcomes fit within this grid and that sort of breaks it up and I can go, Ok I see, what do I need to focus on more in my class to get my kids at the more creative and metacognitive end of the scale?

This is reinforced by several others, with one pointing to its ability to identify ‘holes’ in teaching and another noting that it highlighted the paucity of learning in the classroom that is targeted towards higher order thinking. While another participant acknowledged the potential confusion where an outcome could be classified in a couple of different places, she still maintained, ‘the actual matrix was quite easy to follow and easy to understand and we spent some time discussing that as well when we were discussing what that actually meant on the matrix’.

Given this overall positive reaction to the Instructional Activity Matrix, the less enthusiastic response to its use in the classroom was intriguing. In fact, the interviews revealed that this issue was less related to *Maestro* or the Matrix than the use of learning outcomes themselves to inform teaching. One participant claimed, ‘as teachers we don’t have a lot of time to sit the whole day and look at all the outcomes and break them apart and discuss what they mean’. When interpreted in the light of one of the survey responses which described a preference for identifying intention and success criteria, it is evident that the teachers think of learning in more granular and discrete ways and in a manner particularly targeted towards assessment. For one, the diminished focus on outcomes was because they were busy and tended to just ‘tick them off’ as a requirement, while another contended that the teacher is not able to change the syllabus. For one, the value of classifying outcomes was purely in the planning phase of instruction.

Instead, all participants specifically mentioned assessment as the key area where teachers have influence and this is where the *Maestro* and the Instructional Activity Matrix would be most powerful. One described it in the following terms:

> What that would allow you to do, particularly in my role, I can say to the Year 5 teachers in my stage 'look at your assessments, they're all over at this side of the grid, there's nothing over here that is going to get the kids to analyse, all of your assessments are knowledge based and whatever, I need you to move some of your assessments in this direction', so if it allowed me to do that, I think that would be very useful. I think that as a teacher it would allow you to look at your own assessments and kind of go 'you know what, they're all knowledge based stuff, I really need them to be a bit more challenging and open-ended' so that the kids can show me how much they really know.

This response was typical of several of the interviewees and emphasised the value of *Maestro* for the way in which it would actually be applied in the classroom. Classifying outcomes was an important initial step in testing the system, particularly as these outcomes have been developed through a significant process of expert and peer review and have an inherent authority as curriculum standards. However, these are a limited set that can be pre-classified within the *Virtuoso* system. The real strength of *Maestro* is where teachers can take their own assessments and evaluate the extent to which they address the outcome they are intending to meet.

Limitations

To put the findings of this research into an appropriate context, a number of limitations need to be acknowledged. The small number of participants means that the findings from the survey can be viewed only as trends rather than statistically significant. On the other hand, the triangulation afforded by the interviews allowed for those trends to be interrogated in significant depth. Caution is required in generalising these findings, however, as they represent a cohort of secondary teachers in a single state operating on a national and state-based curriculum. While the comments provided regarding the paucity of outcomes that addressed higher order thinking suggests that the Instructional Activity Matrix can accommodate outcomes pitched at much higher levels such as those defined for undergraduate and postgraduate degrees, further research will need to be undertaken in other settings, particularly in tertiary environments. This is important because the *Virtuoso* platform has international clients and is a scalable enterprise learning management system that is able to be implemented in universities. Finally, further research will need to be undertaken to explore the classification of actual assessment items and how they compare to their stated outcomes.

Conclusion

This study has shown that the Instructional Activity Matrix has the potential to be a useful tool to inform the design and assurance of learning outcomes and assessment. Its strength is undoubtedly in its capacity to prompt consideration of actual practice against intended goals and the potential of an automated system as an efficient prompt for teachers undertaking such activities.

To further enhance the tool, the findings from this study can be integrated into *Maestro* in a number of ways.

1. The current interface is functional but not designed specifically to promote best practice in assessment among teachers. In order for this product to be successful, usability factors will need to be integrated, particularly in terms that highlight its role as a support rather than a turnkey solution. The stated discomfort articulated by a few participants around not knowing the 'right answer' for classifying outcomes means that *Maestro* needs to be presented as a guide that leverages off the expertise of the teacher rather than providing a prescriptive solution.
2. Given most teachers' focus on assessment rather than outcomes, this interface will foreground assessment and use outcomes classification as a means to calibrate that rather than as a goal in itself. This will require further research to assess its validity as an assessment classifications tool as well as providing a workflow for educators to undertake this calibration activity.
3. Outcomes vary greatly between domains and while *Maestro*'s strength is its lack of bias, subject-specific definitions will enhance its capacity to be useful across a range of disciplines. Some curricula such as the NSW English curriculum already provide these while others can be developed via expert review. Teachers will then select the relevant discipline before submitting their assessment for classification and feedback.

These improvements will be undertaken in the next stage of the research and further development of *Maestro*. Given the traction that outcomes-based education has developed within both the secondary and tertiary sectors, and the increased focus on assuring outcomes through valid assessment, there is no doubt that a tool such as this can provide assistance in the design of assessment in a way will be highly beneficial to teachers. This needs, however, to be developed in a manner that acknowledges its role as an objective but imperfect prompt to assist in the more subjective but equally imperfect process of assessment design. Just as the teachers in this study valued the opportunity to develop their understandings of learning outcomes in a socially negotiated space, then *Maestro*'s role will be to provide support for assessment design in a manner that embraces its complexity and supports their expert decision-making, while maintaining the benefits of being an efficient and autonomous performance support system.

References

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Australian Qualifications Framework Council. (2013). *Australian Qualifications Framework Second Edition January 2013*. Retrieved from Adelaide:

Biggs, J. (1989). Towards a Model of School-Based Curriculum Development and Assessment Using the SOLO Taxonomy. *Australian Journal of Education, 33*(2), 151-163.

Birmingham, S. (2015). Higher Education Standards Framework (Threshold Standards) 2015. [Canberra]: Commonwealth of Australia.

Blackboard Inc. (2016). ANZ Blackboard: Reimagine Education. Retrieved from http://anz.blackboard.com/

Bloom, B. S. (1956). *Taxonomy of Educational Objects, Handbook 1: The Cognitive Domain*. New York: David McKay Co Inc.

Bradley, D., Noonan, P., Nugent, H., & Scales, B. (2008). *Review of Australian Higher Education: Final Report*. Retrieved from Canberra:

DEEWR. (2009). Transforming Australia's Higher Education System. Retrieved from ACT, Australia: http://www.innovation.gov.au/highereducation/Documents/TransformingAusHigherED.pdf

Gagne, R. M. (1985). *The Conditions of Learning and Theory of Instruction*. New York: CBS College Publication.

Jonassen, D. H., & Grabowski, B. L. (2012). *Handbook of Individual Differences Learning and Instruction*. New York: Routledge.

Jonassen, D. H., Tessmer, M., & Hannum, W. H. (1999). *Task Analysis Methods for Instructional Design.* Mawwah, New Jersey: Lawrence Erlbaum.

McMahon, M., & Garrett, M. (2015). *A Classification Matrix of Examination Items to Promote Transformative Assessment*. Paper presented at the Global Learn 2015, Berlin, Germany. http://www.editlib.org/p/150861

McMahon, M., & Garrett, M. (2016). *Mind vs. Machine: A comparison between human analysis and autonomous natural language processing in the classification of learning outcomes*. Paper presented at the EdMedia: World Conference on Educational Media and Technology 2016, Vancouver, BC, Canada. https://www.learntechlib.org/p/173011

Merrill, M. D. (1983). Component Display Theory. In C. Reigeluth (Ed.), *Instructional Design Theories and Models: An overview of their current status.* (pp. 279-334). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Thompson-Whiteside, S. (2012). Setting Standards in Australian Higher Education? *Australian Association of Instittutional Research, 17*(1), 27-38.

Please cite as: McMahon, M. & Garrett, M. (2016). Evaluation of a learning outcomes taxonomy to support autonomous classification of instructional activities. In S. Barker, S. Dawson, A. Pardo, & C. Colvin (Eds.), *Show Me The Learning*. *Proceedings ASCILITE 2016 Adelaide* (pp. 421-430).

Note: All published papers are refereed, having undergone a double-blind peer-review process.