How authentic is it? Evaluating the products of an authentic assessment task

Harrison Gray, Kay Colthorpe, Louise Ainscough and Hardy Ernst
School of Biomedical Sciences, University of Queensland, Australia.

Abstract

Authentic assessment tasks resemble activities that are practiced in the workforce. These tasks are valued because they represent what students wish to accomplish as professionals, positively influencing their aspirations and motivations by explicitly demonstrating relevance of assessment tasks. However, given the choice available to students in completing authentic tasks and novelty of outcomes, the products of such assessment may vary in authenticity. This study aimed to develop a method of evaluating authenticity in student assessment products. Second year occupational therapy students (n=59) completed a written factsheet assignment about a disease or condition. The students' products were evaluated for authenticity using a novel rubric developed during the study. The results demonstrate that authenticity in the product of an authentic assessment task is measurable, but varies widely across a cohort, with most products demonstrating moderate to high authenticity. However, there was no correlation between authenticity and course grade. Neither was there a correlation between the grade for this authentic task and a verbal authentic task in another course. These findings suggest that students, at this stage of their education, may not yet have progressed from writing like a professional to acting like one.

Keywords: authentic assessment; factsheet; evaluating authenticity.

1. Introduction

Tertiary institutions across the globe are being encouraged to create authentic curricula to better prepare graduates for the working world (Trede, Macklin & Bridges, 2012). Part of that preparation involves undergoing professional socialisation, to aid understanding of workplace cultures and learning of professional roles (Cornelissen & Van Wyk, 2007). This allows students to acquire knowledge and develop the skills associated with becoming a member of their chosen profession, and adopting its culture, norms and values (Cornelissen & Van Wyk, 2007). An effective way of achieving professional socialisation is through authentic assessment, where individuals experience workplace conditions or tasks that mimic career practises (Burkill, 2009; Kohnen, 2013).

Boud and Falchikov (2006; 2007) define authentic assessment as something that closely resembles activities that are practiced in the workforce, separate from the artificial constructs offered in university courses. Students value authentic assessment because it is a representation of what they would like to accomplish in the workforce (Herrington & Herrington, 1998; Meyers, 2009). It has the potential to positively influence students by raising their aspirations and increasing motivation through explicitly demonstrating the relevance of curriculum activities and career alignment (Frey & Schmidt, 2007).

Many guides on the design of authentic assessment tasks have been published, across various disciplines, and the key features of authentic tasks have been well characterised (Gulikers, Bastiaens & Kirschner, 2006; Burton, 2011; Herrington & Herrington, 1998). These features include: considerations of the fidelity of task to the real world, including realistic conditions and using a variety of resources; that the task produces a polished, valuable product (Gulikers et al, 2006); and the task requires higher order thinking, reflection, metacognition and self-assessment (Herrington & Herrington, 1998; Burton, 2011). Authentic assessment should allow students to determine tasks, make judgments and choices, and generate diverse and novel responses (Burton, 2011). The tasks should be ill-structured, need open-ended inquiry and the construction of novel ideas, and should seamlessly integrate with other assessment (Frey & Schmidt, 2007; Burton, 2011).

However, despite the considerable interest in authentic task design, relatively few studies have evaluated the authenticity of the products students create for such tasks. Given that authentic assessment design should allow students to have choice in their approach to the task and novelty in response, it is likely that students may choose to create products of varying levels of authenticity. The extent to which students create authentic products may be influenced by their knowledge of the skills and practices of their profession, and the tasks they will undertake as professionals. In addition, students are likely to be influenced by their perceptions of the value of the assessment task, particularly its relevance to their

profession, and the instructions and guidance they receive. The aim of this study was to develop a method to evaluate authenticity in the product of an authentic assessment task.

2. Methods

The participants for this study were second year Bachelor of Occupational Therapy (OT) students undertaking a physiology course, 'Human Function in Health and Disease II' at the University of Queensland in semester 2, 2016. Students (n=117) in the course had an average age of 20.5 years, 93% were female and 12% were international students.

Students completed a 'factsheet' assignment, written as if for a healthcare team in a rural hospital. In this authentic assessment task, which was designed to meet the key features of authentic task design (Burton, 2011), students created an information sheet describing a disease related to the physiology in the course, how the disease impacts on occupational performance, and the role of an occupational therapist in its management. The factsheets contributed 17% to the overall course grade, and were marked on scientific content, and on the descriptions of the impact and management of the disease. Students were provided with guidelines for the assignment, however these did not include explicit information regarding layout or writing style except to state that it should be "academic in nature". The assignments were not graded on format or authenticity, so a rubric measuring how authentic each factsheet appeared was created and is described below. Data from consenting students, including their overall course grade and performance in an authentic verbal clinical examination task in a concurrent OT course were also collected and analysed.

Ethics approval was received from the institutional Human Research Ethics Committee prior to the beginning of this study. Of this cohort, 59 students (50%) provided informed consent to participate in the study. A t-test was used to evaluate whether any significant difference existed between the examination performance of students who had consented (44.3 +/- 6.9 out of 60) and the entire cohort (43.3 +/-7.6; p=0.37); as no difference existed it can be assumed that participating students were academically representative of the cohort. Consenting students' data was assigned a de-identified code prior to analysis, which allowed matching of all data sources.

2.1. Measuring authenticity of student product

As no rubrics exist to judge authenticity in student product, a novel rubric was developed (Table 1), using (i) Burton's framework of authentic assessment design (Burton, 2011); (ii) a factsheet checklist developed by the University of North Dakota (2018); and (iii) the assignment criteria rubric. Burton's framework of authentic assessment design presents compelling features of authentic assessment design and 'yes/no' questions derived from each feature. Those features relevant to the product of an authentic task include its fidelity

to the real world, and the judgements and choices students make regarding sub-tasks. The latter is also addressed in the University of North Dakota's factsheet checklist. It is comprised of two main sections: 'content', whether the student has used elements of professionalism in the written content, and 'layout', whether the student has used elements of professionalism in the design. The inclusion of these elements could also be answered with yes/no.

The final compelling feature identified by Burton (2011) relevant to the product of an authentic assessment relates to the accuracy of the work, specifically, whether it is a polished product in its own right. In the context of this product, accuracy was represented by the students' explanations of the subject matter and occupational therapy role, specifically whether these were appropriate and correct. To evaluate this, the marks awarded by examiners to each factsheet on criteria for pathophysiology, symptoms and impact on OT were summed, giving a mark out of 75. To adapt these to the authenticity rubric, marks were stratified into three tiers of equal value: tier 1 = 34-47 marks, tier 2 = 48-61 marks, and tier 3 = 62-75 marks, and a corresponding score of 1-3 allocated for each assignment.

Each factsheet was then graded based on the authenticity rubric (Table 1), giving an overall authenticity score from all criteria out of 9. These were considered as low (1-3), moderate (4-6) and high (7-9) range of authenticity scores. Once constructed, the reliability of the rubric was established via inter-rater reliability, where a naïve researcher independently graded 100% of the factsheets for authenticity using the rubric. Initial agreement between researchers was 95%. Discrepant responses were discussed, and grading revised as appropriate.

To identify if any relationships existed between authenticity in this written task and course grades, a Spearman's rank correlation was performed. These aspects were also compared to students' performance in another authentic assessment task, a verbal clinical examination, that took place in a concurrent OT course. Results are expressed as mean and standard deviation, and were considered significant if $p < 0.05$.

Table 1. Degree of authenticity rubric

Content: **Does the student mimic a professional?**	**Layout:** **Does the student exercise judgment or choice with subtasks?**	**Accuracy:** **Is it a polished product in its own right?**
Has the student used bullet points/tables? Score = /1	Has the student used a template? Score = /1	Pathophysiology Marks = /40
Is their written work self–contained? Score = /1	Has the student incorporated graphics into their assignment? Score = /1	Symptoms Marks = /15
Have they employed language that is accessible for a lay audience throughout their assignment? Score = /1	Does the student include any indication of OT identity? Score = /1	Impact on OT Marks = /20
		Total marks = /75 34-47 marks = Tier 1 48-61 marks = Tier 2 62-75 marks = Tier 3
Content score = /3	Layout score = /3	Accuracy score = /3

3. Results

Overall, students' factsheets (n=59) had an average authenticity score of 5.37+/-1.61 out of 9. The majority of students (66%) incorporated elements of professionalism in their factsheet assignment, so consequently most had moderate authenticity scores ranging from 4-6 (Figure 1). With regards to authentic writing, 51% of students elected to write in a professional voice. In terms of presentation subtasks, 65% of students opted to use a template; images or graphics were included by 72% of students, but only 13% displayed any sort of identification as an OT in their assignment. The majority of students (80%) used bullet points and/or tables, which are also indicative of an authentic product. The marks awarded by examiners on criteria relating to the pathophysiology, symptoms and impact on OT gave a mean accuracy score of 56.42+/-10.36 out of 75.
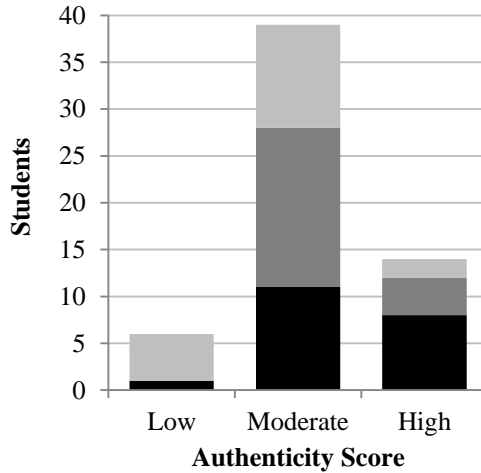
Figure 1: Authenticity scores for factsheet assignments. Each factsheet assignment (n=59) was allocated a score using the authenticity rubric and divided into low moderate and high score ranges. The grey scale represent scores of low (black), middle (dark grey) and high (black) levels within each range.

Authenticity score was not significantly correlated with physiology grade, nor with scores on the OT verbal clincal examination task or grades in the OT course. However, significant positive correlations were observed (Table 2) between scores on the OT task, grades in that course (r=0.79; p<0.0001); and physiology grades (r=0.301; p<0.05). To elucidate if there were any specific aspects of authenticity that were related to grades, further Spearman's correlations were performed. These showed that accuracy was positively correlated with physiology course grade (r=563; p<0.0001), but not other OT task score or course grade, whereas content and layout did not correlate with any of these aspects.

Table 2. Spearman's rank correlation matrix of authenticity, scores and course grades.

	Authenticity	**OT task score**	**OT Grade**
Authenticity	-		
OT task score	0.018		
OT Grade	0.117	0.790****	
Physiology grade	0.198	0.301*	0.580****

* p<0.05; **** p<0.0001

4. Discussion

The aim of this study was to examine the authenticity of the products of assessment created by university students. Before this could be addressed, methods of characterising authenticity needed to be created, which gave rise to a preceding aim: to develop a rubric to evaluate authenticity in student products. The synthesis of existing literature allowed the creation of a novel rubric that could quantify authenticity and enable the identification of discernible differences in the level of authenticity of student assessment products.

The main distinction in students' factsheets was whether they were presented like a student assignment or as a professional product. This requirement is mostly satisfied from the appearance of content and layout, where most students exercised an option to use aesthetic elements such as templates, graphics, bullet points and tables to make their assignment look professional. Of particular interest was whether students decided to write like a professional, in that they used language that was accessible to a lay audience and self-contained, with just over half doing so. The majority elected to employ elements of professionalism even though it was not a requirement of the assessment task.

The authenticity rubric adapted criteria from three different sources: Burton's compelling features of authenticity (2011), The University of North Dakota's factsheet checklist (2018), and the assignment criteria. The need to amalgamate these sources arose from a lack of pre-existing methods for evaluating authenticity in a student product; consequently, it is a new and unverified measure. Limitations are unavoidable when the research conducted is novel, but certain actions should be taken to ensure the measurement is reliable (Cowin, Johnson, Wilson & Borgese, 2013). Thus, an inter-rater reliability test was used (Armstrong, Gosling, Weinman & Marteau, 1997). The resultant agreement suggests that authenticity was judged similarly by a naïve researcher in this measurement system.

Rather than 'checking boxes' in order to generate an authenticity score, potentially an easier way to evaluate authenticity is to observe the subject in an authentic setting in real-time. This occurred during the OT clinical examination task in the concurrent course. Although also authentic, the factsheet is marked on how accurately students describe physiological symptoms and intervention strategies in their writing, whereas the clinical examination is marked largely on verbal interaction with a patient. While authenticity was not marked explicitly in the OT task, it is not unreasonable to suggest that examiners marking clinical assessments are discretely marking for authenticity overall, as they are looking for similarities between the student's actions and that of health professionals (Thomas, Saroyan & Dauphinee, 2011).

The lack of a significant relationship between authenticity in the physiology and OT tasks imply a distinction between the authentic assessment types, as the correlation between physiology and OT course grades indicate students should achieve similar results across

both courses. It is possible that at this point in their program, students are only at the stage of writing like an OT, rather than acting like one (Rodger, Turpin & O'Brien, 2015). This is supported by the fact that few students assumed the role of an OT in their factsheet. It seems clear that their professional writing has not yet progressed to seeing themselves as medical professionals, which is directly related to professional socialisation (Smith & Hatmaker, 2009). For this reason, it would be beneficial to evaluate the products of authentic tasks from later in the students' program, to evaluate the progress of the students' self-perceptions.

At the time of these courses, the OT students were only just beginning their clinical placements, so had not yet had much contact with practising occupational therapists through their program. Authentic settings, such as placements, have been shown to present major challenges for health students (Bramming, 2007). The strongest learning on the path to professionalism occurs when students face a crisis that is challenging and confronting, which forces them to stop, think and reflect (Bramming, 2007). The fact that these students have not yet faced such a challenge is reflected in their inconsistent grades for authentic assessment tasks in physiology and OT. Potentially, when these methods of socialisation occur, students will progress from writing like a professional to acting like one.

References

Armstrong, D., Gosling, A., Weinman, J., & Marteau. T. (1997). The place of inter-rater reliability in qualitative research: an empirical study. *Sociology*, 31(3), 597-606.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long- term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399-413.

Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education: Learning for the longer term,* New York, Routledge.

Bramming, P. (2007). An argument for strong learning in higher education. *Quality in Higher Education,* 13**,** 45-56.

Burkill, S., Dunne, L., Filer, T. & Zanstra, R. (2009). Authentic voices: Collaborating with students in refining assessment practices. In *ATN Assessment Conference 2009: Assessment in Different Dimensions.* 84-93.

Burton, K. (2011). A framework for determining the authenticity of assessment tasks: applied to an example in law. *Journal of Learning Design,* 4**,** 20-28.

Cornelissen, J. J., & Van Wyk, A. S 2007. Professional socialisation: An influence on professional development and role definition. *South African Journal of Higher Education,* 21**,** 826-41.

Cowin, L. S., Johnson, M., Wilson, I. & Borgese, K. (2013). The psychometric properties of five Professional Identity measures in a sample of nursing students. *Nurse Education Today,* 33**,** 608-613.

Frey, B., & Schmidt, V. (2007). Coming to terms with classroom assessment. *Journal of Advanced Academics,* 18**,** 402-423.

Gulikers, J., Bastiaens, T. & Kirschner, P. (2006). Authentic assessment, student teacher perceptions: the practical value of the five-dimensional framework. *Journal of Vocational Education and Training,* 58**,** 337-357.

Herrington, J. H., & Herrington, A. (1998). Authentic assessment and multimedia: how university students respond to a model of authentic assessment. *Higher Education Research and Development,* 17**,** 305-322.

Kohnen, A. M. (2013). The Authenticity Spectrum: The Case of a Science Journalism Writing Project. *English Journal,* 102**,** 28–34.

Meyers, N. M., & Nulty, D. D (2009). How to use (five) curriculum design principles to align authentic learning environments, assessment, students' approaches to thinking and learning outcomes. *Assessment & Evaluation in Higher Education,* 34**,** 565-577.

Rodger, S., Turpin, M., & O'Brien, M. (2015). Experiences of academic staff in using threshold concepts within a reformed curriculum. *Studies in Higher Education*, 40(4), 545-560.

Smith, A. E., & Hatmaker, D.M. (2009). Knowing, doing and becoming: professional identity construction among public affairs doctoral students *Journal of Public Affairs Education,* 20**,** 545-564.

Thomas, A., Saroyan, A., & Dauphinee, W. D. (2011). Evidence-based practice: a review of theoretical assumptions and effectiveness of teaching and assessment interventions in health professions. *Advances in Health Sciences Education*, 16(2), 253-276.

Trede, F., Macklin, R., Bridges, D. (2012). Professional identity development: a review of the high education literature. *Studies in Higher Education,* 37**,** 365-384.

University of North Dakota. (2018). Factsheet checklist. Retrieved from https://ruralhealth.und.edu/communication/factsheets [Accessed 10/12/2018]