



Changing the Quality of Teachers' Written Tests by implementing an Authentic Assessment Teachers' Training Program

Verónica Villarroel

Dr., CIME, Faculty of Psychology, Universidad del Desarrollo, Chile, vvillarroel@udd.cl

Daniela Bruna

Dr., CIME, Faculty of Psychology, Universidad del Desarrollo, Chile, dbrunaj@udd.cl

Gavin T. L. Brown

Prof., The University of Auckland, New Zealand, gt.brown@auckland.ac.nz

Claudio Bustos

Dr., Faculty of Medicine. Universidad de Concepción, Chile, clbustos@udec.cl

This case study aimed to change the construction of teachers' written tests so that items were designed to assess competencies in an authentic and challenging way. A small group of five psychology teachers participated in 10 sessions of an authentic assessment faculty-training program, to learn to assess problem-solving competencies for situations typically faced by professionals in the workplace. The authentic assessment training emphasized the incorporation of three main characteristics: (1) inclusion of a realistic context, (2) measurement of higher order thinking skills, and (3) development of evaluative judgment, concerning the quality of their own performance. Post-training the items' construction was analysed, according to their type and authenticity. Mixed effects logistic regression showed a statistically significant increase in open-response items, and two-way ANOVA indicated that cognitive challenge improved. The results showed written tests had: a) more open-response items of higher cognitive complexity, b) fewer items requiring closed, memorized responses, c) more use of realistic contexts to measure knowledge in a situated way, and d) improvement in curriculum alignment of tests and greater consistency in measuring competencies. These advances validate the authentic-assessment training program for a better written assessment design.

Keywords: assessment, authentic assessment, authenticity, psychological skills, teachers

INTRODUCTION

The number of higher education graduates is progressively increasing. The participation of students in higher education is expected to reach more than 700 million people by the year 2100, representing a tenfold increase from 1970 (Roser and Nagdy, 2018).

Citation: Villarroel, V., Bruna, D., Brown, G. T.L., & Bustos, C. (2021). Changing the Quality of Teachers' Written Tests by implementing an Authentic Assessment Teachers' Training Program. *International Journal of Instruction*, 14(2), 987-1000.

Universities face the challenge of graduating professionals recognized as competent, chosen by future employers and valued by their workplace.

To achieve this, most undergraduate programs have graduation profiles that guide their educational processes, ensuring the achievement of the skills necessary for good professional performance in a specific discipline. Therefore, the skills selected for the profile correspond to the needs of the profession and the context in which they are performed (Davies, Mangan, Hughes & Slack, 2013; Tholen, James, Warhurst & Commander, 2016). Professional competencies require mastery of technical knowledge and generic and specific skills, as well as higher-order thinking abilities (Guzzoni, Male & Miller, 2015; Medland, 2016).

Employability is defined as the ability to find, create and sustain meaningful work across the career lifespan, having the skills, knowledge, understanding and personal attributes that make a person more likely to choose and secure occupations in which they can be satisfied and successful. So, employability must focus on abilities and must integrate the metacognitive capacities to continue learning (Bennett, 2018). Precisely, some of the most sought skills by employers are the ability to solve problems and think critically (Oliveri and Markle, 2017; Partnership for 21st Century Skills, 2010; Wiggins, Hammar, Larsson, Pauli, & Worrell, 2016); such learning outcomes are also valued according to the Guidelines for the Degree in Psychology (APA, 2017).

How can the achievement of the competences required for graduating from a program be ensured? Assessment of learning is the axis of the educational process, and fulfils the role of aligning learning outcomes to the learning activities offered by teaching and students' achievement (Biggs, 2003). Assessment has powerful backwash effects on all teaching and learning activities (Watkins, Dahlin, & Ekholm, 2005), has a strong influence on what and how students' study and learn (Kearney, Perkins & Kennedy-Clark, 2015), and influences students' learning and future employability (Ajjawi, Tai, Nghia, Boud, Johnson, & Patrick, 2020).

A major challenge for professors is the ability to design valid assessment methods that really measure students' higher order thinking skills, which are described in undergraduate psychology programs' graduate profiles, and included in professional standards of different countries (Australian Psychological Society, 2013; Markle, 2017, Oliveri & Markle, 2017). Such assessments need to be ecologically valid, clearly reflecting the content and context of professional practice, while requiring the demonstration of complex thinking skills (Falchikov, 2004; Nicholson Perry, Donovan, Knight, & Shires, 2017).

It is often thought that the only way to assess high-order skills is through performance-based tasks, such as problem-based learning, role-playing, portfolios, or daily reports (Duda, Susilo, & Newcombe, 2019). These kinds of instruments are crucial to provide information of students' ability to transfer knowledge to real contexts, as well as being more authentic and motivating (Zaim, Refnaldi, & Arsyad, 2020). Nevertheless, more than 70% of undergraduate programs assess leaning through written tests (Gitanjali, 2016; Pfund, Norcross, Hailstorks, Stamm, & Christidis, 2018). These tests, while trying to cover the whole range of learning outcomes, tend to focus on accumulation of

information, partial understanding of content, and not promote deep approaches to learning (Biggs, & Tang, 2011; Price, Carroll, O'Donovan, & Rust, 2011; Endedijk, & Vermunt, 2013). Most items of these kinds of tests reward memory, which can lead to rapid forgetting (Rawson, Dunlosky, & Sciarrelli, 2013).

Thus, an important mechanism for ensuring higher-order thinking is to modify the kind of items used in written tests to assess student learning, making it more authentic and challenging (Duncan & Buskirk-Cohen, 2011; Endedijk, & Vermunt, 2013). The biggest obstacle in creating tests that assess higher-order thinking skills is that teachers are not well-trained in designing written tests that require the use of complex thinking (Deneen & Boud, 2014; Medland, 2016, Yläne and Nevgi, 2007). Such tests would allow students to gain a deeper understanding of the content (Jensen, McDaniel, Woodard, & Kummer, 2014), and show better stability in remembering what was learned over time (Rawson, et al., 2013). Fundamentally, university instructors can be world-class in their discipline, without knowing how to write appropriate tests and examinations. The main goal of this paper is to demonstrate the positive change in written test construction by applying an authentic faculty training program. In addition, to show that it is possible to assess competencies in an authentic and challenging way through written tests.

Authentic Assessment Movement.

Authentic assessment is an approach that assures the achievement of in-depth and quality learning in contrast to standardized memory-focused items (Martinez, O'Brien, Roberts, & Whyte, 2018). The idea of authenticity in assessment is that students use knowledge to show effective and creative performance, achieving a situated learning that mimics the complexity and contradictions that students are likely to face in the real world (Kearney, 2013; Munandar, Maryani, Rohmat, & Ruhimat, 2020; Saye, 2013). Authentic assessment ensures that all students are given the opportunity to show what they are capable of while giving teachers the necessary information for establishing a balanced and fair assessment to each student (Hanifah & Irambona, 2019).

Authentic assessment reduces the gap between what is learned in university and what is required in the external world (Gulikens, Bastiaens, Kirschner, 2004; Neely & Tucker, 2012). It seeks to engage students with problems or important questions, which are worthwhile beyond the classroom, so that assessment tasks become replicas or analogues of problems faced in working life (Raymond, Homer, Smith & Gray, 2012). This is especially important in vocational training. These students really need to practice and show performance, but it is still so for the rest of the careers, since in all of them it is required the development of deep learning, and to apply and transfer knowledge to other contexts.

In relation to its benefits, authentic assessment develops higher-order cognitive skills (Ashford-Rowe, Herrington, & Brown, 2014), prepares test-takers for autonomous practice (Carter, Sidebotham, Creeedy, Fenwick, & Gamble, 2015), and improves academic engagement (Kearney & Perkins, 2014), motivation for the learning process (Nicol, Thompson & Breslin, 2014) and self-regulation capacities (Ling Lau, 2013).

In previous higher education studies in Chile, an authentic assessment model has established three core dimensions, which constitute the assessment process: a) realism, b) cognitive challenge, and c) evaluative judgement (Villarroel, Bloxham, Bruna, Bruna, & Herrera-Seda, 2018). Realism refers to contextualizing and situating the assessment in professional work-settings, so students can use what they have learned to provide an answer that allows them to value knowledge as a means of understanding and solving problems. Cognitive challenge refers to the need to measure higher order cognitive abilities, mobilizing students to use and transfer their knowledge. Finally, the aim of evaluative judgement is that students develop and incorporate quality criteria so they can judge their own, and their peers' work, reflecting on the improvements that they might have (Villarroel, Boud, Bloxham, Bruna, & Bruna, 2020).

Teaching training program in higher education.

Enriching teachers' pedagogical skills through higher education training courses is a need that has grown over time (Postareff, Lindblom-Ylänne & Nevgi, 2007). The most complex area to transform in teaching practice is the assessment of learning; frequently considered the Achilles' heel of education (Medland, 2016). University teachers receive less training in assessment design and show more resistance to change (Brush & Saye, 2008; Deneen & Boud, 2014; Pereira and Flores, 2016). For example, statistical analysis of two 50-item multiple-choice tests created by instructors of a course called "How People Learn" showed that half of the items had poor distractors, inverse discrimination, or high guessing parameters (Brown & Abdulnabi, 2017). The assessment and feedback practices of clinical psychology supervisors in the last year of psychology education have important deficiencies (Gonsalvez, Wahnnon & Deane, 2017) that could be redressed by using authentic assessment, including observation techniques, formative assessment, and dialogue and oral communication in the feedback processes.

This article seeks to promote the use of authentic assessment in psychology by providing concrete examples of how teachers can transform written test items to make them more realistic, contextualized, and cognitively challenging, following the principles of the authentic assessment model, on which they were trained. The research question is related to the magnitude of the change between the items of the written tests before and after the training. It is interesting to know if teachers manage to change, how they change, and if they can apply the principles of authentic assessment when they design their assessments.

METHOD

Teacher training related to authentic assessment was carried out with the aim of changing written test characteristics. Using quantitative data analysis, change in test construction was measured before and after teacher training in authentic assessment methodology. This allowed the analysis of repeated measures with the same sample of teachers.

Participants

Five professors from the Faculty of Psychology of Universidad del Desarrollo in Chile volunteered for training. All teachers, two women and three men, had a master's degree

with an average teaching experience of 18 years, consistent with an average age of 46 years old. The teachers taught third- and fourth-year undergraduate students.

Procedure and characteristics of the intervention

After receiving information on the objectives of the study, the 5 teachers agreed to participate. All of them were asked to submit written tests that they had previously applied to their students, and thus, there was a baseline regarding the type of items they built to measure learning.

The teachers participated in a 60-hour training program (40 in-person and 20 independent study) during a semester. The ten face-to-face sessions, each lasting four hours, were structured as workshops, with a constructivist and collaborative approach, to build knowledge and practice skills. Teachers worked on the principles of authentic assessment (i.e., realism, cognitive challenge, and evaluative judgment) to transform test items into a more open-ended, analytical, and contextualized format.

Realism can be accomplished by presenting a real context that describes and delivers a frame in which a problem is to be solved. Items which have rich context simulating real-work situations that function as a proxy for professional performance can be drafted, even when the course itself is not intended for the direct preparation of a profession (Bosco & Ferns, 2014). In authentic assessment, the selection of the context that will be assessed constitutes what really matters. To facilitate this adequate selection, we proposed three sources: the graduate profile (Hart, Hammer, Collins & Chardon, 2011), course learning outcomes, and work requirements (Maxwell, 2012).

Authentic assessment expects students to employ higher-order cognitive skills related to using, modifying, or rebuilding knowledge into something new (Thornburn, 2008). It is possible to group cognitive abilities in three blocks (Anderson & Krathwohl, 2001). The first is related to memory skills, such as recognition or understanding of information; the second involves cognitive skills for information management (to compare, analyse, relate, calculate, interpret); and the third comprises performance (to judge, decide, criticize, suggest, design, innovate, propose or invent). Authentic assessment privileges the measurement of cognitive abilities at the second and third levels, which are oriented to "why" students learn the content they learn (Avery, Freeman, & Carmichael, 2012).

Furthermore, in authentic assessment it is necessary to include activities in which students develop an evaluative judgment about their own performance, such as occurs at work (Kearney et al., 2015). Three strategies are proposed to develop this capacity: a) having students develop marking criteria which engages them in a deep understanding of knowledge, because they must review and look for information to create marking guidelines (O'Donovan, Price, & Rust, 2008), b) engaging students in peer review to develop their ability to judge others' work, such as occurs in workplace, and, c) using self-assessment in judging and defending their own work, which involves them in gathering and responding to feedback (Boud & Molloy 2013), in order to improve their performance (Lipnevich, Berg, & Smith, 2016).

Participants had to apply these principles to build new tests during the 10 weekly workshops, received feedback that corrected their work, and were advised on how to

improve. A strength of this training is that it was guided by the principles of authentic assessment (i.e., realism, cognitive challenge, and evaluative judgment). These guidelines were applied during the training sessions. For example, when teachers worked on transforming their own tests, which they had created before training, they faced tasks that involved evaluating, criticizing, designing, and proposing, which denoted high cognitive complexity. They also had to deliver feedback to their peers and self-evaluate their own test item writing based on established quality criteria, using their evaluative judgment. In other words, participants had to use these same principles to build their own authentic items.

Data analysis

To assess the changes generated by the training, the test items created before the training were analyzed, according to their type (open or closed question) and authenticity indicators such as: realism (presence and quality of context); cognitive challenge (three levels: memory, analytical, or transfer) and alignment with specific and generic competencies of the graduate profile in psychology. Each item was analyzed by two different judges, and the correlation between them was $r = .88$, suggesting sufficient agreement. The analysis of these tests informed the initial feedback given each participant. This procedure was repeated with the post-intervention test created by trained participants. A variety of univariate techniques (e.g., ordinal regression and one-way analysis of variance) were performed between pre- and post-intervention tests for each of the four dimensions.

In the first analysis, logistic regression was used to account for change in the percentage of the dichotomous item type variable. The percentage score was used because the number of item types was not the same before and after the training. One-way analysis of variance (ANOVA) was used to compare the average of the score for each polytomously-scored rubric indicator (i.e., realism, cognitive challenge, alignment with specific competences and alignment with generic competences).

FINDINGS

Change in type of items after authentic assessment training

A total of 708 ($n = 414$ pre-intervention, $n = 294$ post-intervention) items were classified dichotomously according to their format: open (e.g., case analysis, short written response, long written response and problem solving) or closed (e.g., multiple choice or true-false). Two teachers were excluded in analysis because they did not use closed items in the pre-intervention tests. Hence, the item type data came from three teachers at two times. Before the intervention, the ratio of open to closed was 41:59 and after the course, it was 51:49. Table 1 provides the results for a mixed effects logistic regression test of time and teacher. The likelihood ratio test showed the model was statistically significant ($\chi^2(3) = 24, p < 0.001$). The only significant coefficient was the effect associated with Time 2 after the intervention (OR = 10.91, 95% CI = 3.41, 60.57; $d = 1.32$). This shows a statistically significant and substantial increment in open format items by post-intervention. The Hosmer and Lemeshow goodness of fit test was non-significant, $\chi^2_{(8)} = 9.78, p = 0.28$, so we can assume that the model was correctly specified.

Table 1
Logistic regression for the change in the type of items

| Coefficient | Estimate | se | <i>p</i> |
|--------------------|----------|------|----------|
| Constant | -0.46 | 0.36 | 0.20 |
| Teacher= Teacher 2 | -1.09 | 0.69 | 0.11 |
| Year= 2 | 2.39 | 0.68 | <0.001 |
| Teacher X Year 2 | -1.21 | 0.96 | 0.21 |
| σ^2 Test | 0.868 | | |

Note: Constant represents effect for teacher 1 in the first year (before the intervention).

Change in authenticity after assessment training.

To assess changes in the four authenticity dimensions, two-way ANOVA was performed on each dimension, using teacher and time as main effects, and interaction between teacher and time as an indicator of individual differences in the implementation (Table 2). Only for cognitive challenge was there a significant interaction effect between teacher and time, indicating that changes in cognitive challenge were specific to individual teachers, and were not common across this small sample. For two of the dimensions (i.e., alignment with specific competencies and realism) there were large main effects for time ($\eta^2_{\text{partial}} > .26$), while the effect for alignment with generic competencies was small. Simultaneously, there were large differences among the teachers for all three dimensions. Hence, the qualities of tests improved after the workshops but the effect was variable among the participants. In the analysis of the ANOVA assumptions, regression linearity is observed in all the cases, as well as variance homogeneity in all time and teacher combinations using the Levene test. Normality of the residuals was found for cognitive challenge, but not for the rest of the variables. There was a mild degree of positive asymmetry in the three variables, in the range of 0.81- 0.84, thus, it was considered unnecessary to perform more adjustments.

Table 2
Interaction effects between teacher and year for authenticity dimensions

| Dimension | Complete model (<i>df</i> =7) | | Time (<i>df</i> = 1) | | Teacher (<i>df</i> =5) | | Time X Teacher (<i>df</i> =1) | |
|---|-----------------------------------|-----------------------|-----------------------|----------|-------------------------|----------|-----------------------------------|----------|
| | <i>F</i> | <i>R</i> ² | <i>F</i> | η^2 | <i>F</i> | η^2 | <i>F</i> | η^2 |
| Cognitive Challenge | 2.326 * | .08 | 1.14 | 0.03 | 1.07 | 0.09 | 4.07* | 0.27 |
| Alignment with specific competencies | 6.449** | .48 | 16.73** | 0.34 | 9.2** | 0.46 | 2.42 | 0.18 |
| Alignment with generic competencies | 2.801* | .36 | 4.15* | 0.11 | 5.66** | 0.34 | 0.25 | 0.02 |
| Realism | 6.865** | .52 | 18.62** | 0.36 | 10.21** | 0.48 | 2.07 | 0.16 |

Notes: * = $p < 0.05$, ** = $p < 0.01$. ANOVA uses type II sum of squares. *df* for residuals = 33.

Examples of the transformation of the authenticity's dimensions in the tests

It seems useful to illustrate the kinds of changes observed in this study in terms of realism, cognitive challenge and evaluative judgment in the design, application, and feedback process of their written tests. Examples are taken from real items drafted by the participants and illustrate differences in realism and cognitive complexity before and after the training program.

Incorporating realism in the context design of the test

Professor A. In a classical psychological theories test, related to behaviourism, Example 1 illustrates the relevance of building a realistic context.

Pre-training:

Explain two advantages and two disadvantages of operant conditioning theory.

Post- training:

The government is evaluating giving an economic bonus to women with the aim of increasing the birth rate in the country. This would be given after the birth of their third child, and could increase until the fifth child arrives. In relation to this measure, answer the following questions:

- a) If the strategy was effective, characterize the function of the bonus in the increment of the country's birth rate, according to the theory of learning identified in the case.
- b) If this strategy was not effective, raise two arguments that would explain this result, considering the case and the principles of the learning theory in which it's based.

Professor B. In a learning processes test, related to learning approaches and theories, Example 2, likewise illustrates the relevance of building a realistic context.

Pre-training:

Explain how students learn using the constructivist approach.

Post- training:

A primary teacher has decided to make changes in the way mathematics is taught. He has introduced the following strategies: i) at the moment of starting with a new content, he makes a conceptual map where he relates the new concepts with others already known, in a subordinate and superordinate way, according to their hierarchy, and ii) uses abacuses, cubes and Lego bricks so students can solve mathematical problems by manipulating materials.

- a) Based on the changes made, infer two of the teacher's epistemological assumptions regarding how students learn.
- b) Relate each strategy used by the teacher with a different theory of learning. For each of them, point out its author and a key concept of the theory.
- c) Suggest a third strategy to incorporate for adolescent students. Justify your answer in relation to the learning approach that sustains the strategy and the abilities of the students of that age.

Assessing complex thinking

Professor C. In a developmental psychology test, related to the concept of temperament. Example 3 shows a transformation in cognitive complexity.

Pre-training:

Define temperament and list the areas that must be assessed to conclude an infant's temperament.

Post- training:

You must evaluate two psychology students in practice, who are required to inquire about children temperament by interviewing the mother in a mirror room. When you evaluate the students, you rate one in the outstanding performance range, and the other in a deficient one. Justify your assessment, giving examples of the type of questions that each student may have asked the mother to investigate the temperament of the infant and which led to your evaluation.

Professor D. Similarly, Example 4 shows in a psychoanalysis test, related to the *Lapsus Linguae* concept, a change in cognitive complexity.

Pre-training:

A "*Lapsus linguae*" can be explained as:

- a) An intrusion of the id into the ego, according to the first topic.
- b) A discharge by association of similarity in the economic model.
- c) An autoplasmic adaptation, according to Anna Freud.
- d) A regression to the service of the ego, according to Sigmund Freud

Post- training:

Susan and Peter were reunited after two months, after the holidays. Upon arriving at the coffee shop, Susan is surprised at Peter's weight gain, but says nothing. Peter says that he went to meet his father in Madrid, whom he had not seen for a couple of years. Susan knew that her friend did not have a good relationship with her father, and she asks him: "How was the encounter with your fat... fat... fat... sorry, father?" Susan blushed. How would you explain Susan's mistake?

- a) An intrusion of the id into the ego, according to the first topic.
- b) A discharge by association of similarity in the economic model.
- c) An autoplasmic adaptation, according to Anna Freud.
- d) A regression to the service of the ego, according to Sigmund Freud.

DISCUSSION

Universities around the world are trying to innovate in the teaching and learning process, however, transforming learning assessment remains the most complex and difficult challenge to tackle. Even though the literature on evaluation indicates the need to evaluate learning in a more authentic way, written tests are the most frequently used and which seem to remain conventional and limited in realism, cognitive demand, and evaluative judgment.

Frequently, items on written tests measure the decontextualized memorization of knowledge as was seen in the pre-intervention baseline tests examined in this study. The negative impact of this trend is that psychology students will focus on remembering theories and only approach practice later in their study or employment. This study demonstrates that the psychology teachers' can be taught the principles of authentic assessment methodology in written tests, creating items with more realistic contexts, in which students have to analyse a problem, and to assess higher-order thinking skills. It is also a clear implication that without training, higher education academics are unlikely to create more effective and realistic test items in their written tests.

Even though, this case study is limited to a very small number of participants, the large effects suggest that extending the training to other disciplines and institutions should result in positive changes in the quality of written testing. The voluntary nature of the sample introduces a bias due to the enthusiasm of participants. However, this represents the reality of most teachers training courses in higher education, which are attended by only those most interested in improving their skills.

Future research could be oriented towards working with larger and heterogeneous groups, making comparisons between teachers from different disciplines and at different levels of academic training (e.g. early career vs. experienced staff; first vs. final year students), contrasting the magnitude of the change in the learning assessment system with teachers' years of experience, their qualifications and their beliefs regarding learning, as well as analyzing the influence of authentic assessment on students' own perception of learning and their concrete results.

CONCLUSION

The change in the area of learning assessment is very challenging, because some teachers are reluctant to transform their practices and instruments, and because some institutions do not offer the conditions to support academic skills in writing test items and tasks that are realistic and require complex cognition. There is also resistance on the part of the students themselves, who from their academic experience have concluded that learning is memorizing, and since they have been successful in conventional methods, they do not want to threaten their performance with innovations in assessment (Struyven & Devesa, 2016).

The main objective was to change the construction of teachers' written tests to achieve design items to assess competencies in an authentic and challenging way, and that was achieved. This training study shows that it is possible to make changes within the testing and examination practices of the classroom, especially when the training is of a practical nature, allowing application of what has been learned and receiving feedback on how to improve.

Based on teacher training, it was possible for teachers to design written tests with more open-response items of higher cognitive complexity, fewer items requiring closed, memorized responses, more use of realistic contexts to measure knowledge in a situated way, improve the curriculum alignment of tests, and create greater consistency in measuring competencies.

REFERENCES

- American Psychological Association (2017). *APA guidelines for undergraduate psychology major*. Washington, DC: Author.
- Ajjawi, R., Tai, J., Nghia, T., Boud, D., Johnson, L., & Patrick, C. (2020). Aligning assessment with the needs of work-integrated learning: the challenge of authentic assessment in a complex context, *Assessment & Evaluation in Higher Education*, 45:2, 304-316, doi: 10.1080/02602938.2019.1639613

- Anderson, L., & Krathwohl, D. A. (2001). *Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Addison Wesley Longman.
- Ali, N., Rose, S., & Ahmed, L. (2015). Psychology students' perceptions of and engagement with feedback as a function of year of study. *Assessment & Evaluation in Higher Education*, 40(4), 574-586. doi:10.1080/02602938.2014.936355.
- Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205-222. doi: 10.1080/02602938.2013.819566.
- Avery, P.G., Freeman, C. & Carmichael, D. (2012). Developing Authentic Instruction in the Social Studies. *Journal of Research in Education*, 12(1), 50-56.
- Bennett, D. (2018). Graduate employability and higher education: Past, present and future. *HERDSA Review of Higher Education*, 5, 31-61.
- Biggs, J. and Tang, C., (2011). *Teaching for quality learning at university: What the student does*, Maidenhead, Berkshire: Open University Press
- Bosco, A.M. y Ferns, S. (2014). Embedding of authentic assessment in work- integrated learning curriculum. *Asia- Pacific Journal of Cooperative Education*. 15(4), 281- 290.
- Boud, D. & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698-712. doi: 10.1080/02602938.2012.691462.
- Brown, G. T. L., & Abdulnabi, H. (2017). Evaluating the quality of higher education instructor-constructed multiple-choice tests: Impact on student grades. *Frontiers in Education*, 2(24). doi:doi:10.3389/educ.2017.00024
- Carter, A., Sidebotham, M., Creedy, D., Fenwick, J. & Gamble, J. (2015). Strengthening partnership: the involvement of health care providers in the evaluation of authentic assessment within midwifery undergraduate education. *Nurse Education in Practice*, 15(4), 327-32. doi: 10.1016/j.nepr.2015.01.013.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Davies, P., Mangan, J., Hughes, A., & Slack, K. (2013). Labour market motivation and undergraduates' choice of degree subjects. *British Educational Research Journal*, 39(2), 361-382. doi: 10.1080/01411926.2011.646239.
- Deneen, C., & Boud, D. (2014). Patterns of resistance in managing assessment change. *Assessment & Evaluation in Higher Education*, 39(5), 577-591. doi:10.1080/02602938.2013.859654
- Duda, H. J., Susilo, H. & Newcombe, P. (2019). Enhancing Different Ethnicity Science Process Skills: Problem-Based Learning through Practicum and Authentic Assessment. *International Journal of Instruction*, 12(1), 1207-1222. <https://doi.org/10.29333/iji.2019.12177a>
- Duncan, T. & Buskirk-Cohen, A. (2011). Exploring Learner-Centered Assessment: A Cross-Disciplinary Approach. *International Journal of Teaching and Learning in Higher Education*, 23(2), 246-259.

- Endedijk, M. D., & Vermunt, J. D. (2013). Relations between student teachers' learning patterns and their concrete learning activities. *Studies in educational evaluation*, 39(1), 56-65. doi: 10.1016/j.stueduc.2012.10.001.
- Falchikov, N. (2004). Involving students in assessment, *Psychology Learning and Teaching*, 3(2), 102-108.
- Gitanjali, M. (2016). The three Rs of written assessment: The JIPMER experience. *Journal of Pharmacology and Pharmacotherapeutics*, 7(3), 115-119. doi: 10.4103/0976-500X.189650.
- Gonsalvez, C. J., Wahnnon, T., & Deane, F. P. (2017). Goal-setting, feedback, and assessment practices reported by Australian clinical supervisors. *Australian Psychologist*, 52(1), 21-30. <https://doi.org/10.1111/ap.12175>
- Gulikers, J., Bastiaens, T. & Kirschner, P. (2004). A five-dimensional framework for authentic assessment, *Educational Technology Research and Development*, 52(3), 67-86. doi: 10.1007/BF02504676.
- Guzzomi, A. Male, S. & Miller, K. (2015). Students' responses to authentic assessment designed to develop commitment to performing at their best. *European Journal of Engineering Education*, 42(3), 1-22. doi: 10.1080/03043797.2015.1121465.
- Hart, C., Hammer, S., Collins, P. & Chardon, T. (2011). The real deal: Using authentic assessment to promote student engagement in the first and second years of a regional program. *Legal Education Review*, 21(1), 97-121.
- Hanifah, M. &, Irambona, A. (2019). Authentic assessment: Evaluation and its application in science learning. *Psychology, Evaluation, and Technology in Educational Research*, 1(2), 81-94, doi: <http://dx.doi.org/10.33292/petier.v1i2.4>
- Jensen, J., McDaniel, M., Woodard, S., & Kummer, T. (2014). Teaching to the test... or testing to teach: exams requiring higher order thinking skills encourage greater conceptual understanding. *Educ. Psychol. Rev*, 26, 306-329. doi: 10.1007/s10648-013-9248-9
- Kearney, S. (2013). Improving engagement: the use of authentic self- and peer-assessment for learning' to enhance the student learning experience. *Assessment & Evaluation in Higher Education*, 38(7), 875- 891. doi:10.1080/02602938.2012.751963
- Kearney, S., & Perkins, T. (2014). Engaging Students through assessment: The Success and Limitations of the ASPAL (Authentic Self and Peer-assessment For Learning) Model. *Journal of University Teaching and Learning Practice*, 11(3), 1-13.
- Kearney, S., Perkins, T., and Kennedy-Clark, S. (2015). Using self- and peer-assessments for summative purposes: analysing the relative validity of the AASL (Authentic Assessment for Sustainable Learning) model. *Assessment & Evaluation in Higher Education*, 41(6), 1-14, doi: 10.1080/02602938.2015.1039484
- Ling Lau, K. (2013). Chinese language teachers' perception and implementation of self-regulated learning-based instruction. *Teacher and Teaching Education*, 31, 56-66. doi: 10.1016/j.tate.2012.12.001

- Lipnevich, A. A., Berg, D. A. & Smith, J. K. (2016). Toward a model of student response to feedback. *Handbook of Human and Social Conditions in Assessment*, 169-185.
- Martinez, M., O'Brien, M., Roberts, K., Whyte, D. (2018), Critical Pedagogy an assessment in higher education: The ideal of "authenticity" in learning. *Active Learning in Higher Education*, 19(1) 9-21 doi:10.1177/1469787417723244.
- Markle, R. (2017). Continuing a culture of evidence: student-level assessment. *ETS research Report*, 10-17.
- Maxwell, T. (2012). Assessment in higher education in the professions: action research as an authentic assessment task. *Teaching In Higher Education*, 17(6), 686-696. doi:10.1080/13562517.2012.725220
- Medland, E. (2016). Assessment in higher education: drivers, barriers and directions for change in the UK, *Assessment & Evaluation in Higher Education*, 41(1), 81-96. doi:10.1080/02602938.2014.982072
- Munandar, A., Maryani, E., Rohmat, D., & Ruhimat, M. (2020). Establishing the Professionalism of Geography Teacher through Authentic Assessment Field Study. *International Journal of Instruction*, 13(2), 797-818. doi.org/10.29333/iji.2020.13254a
- Neely, P. & Tucker, J. (2012). Using business simulations as authentic assessment tools. *American Journal of Business Education*, 5(4), 449-456.
- Nicol, D., Thomson, A. & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102-122. doi:10.1080/02602938.2013.795518.
- Nicholson Perry, K., Donovan, M., Knight, R, & Shires, A. (2017). Addressing professional competency problems in clinical psychology trainees. *Australian Psychologist*, 52, 121-129. doi:10.1111/ap.12268
- Oliveri, M.E., & Markle, R. (2017). Continuing a culture of evidence: expanding skills in higher education, ETS Research Report, Princeton, NJ.
- Partnership for 21st Century Skills (2010). *American Management Association Critical Skills Survey*. Survey, Tucson: P21.
- Pereira, D., and Flores, M. A (2012). Percepções dos estudantes universitários sobre avaliação das aprendizagens: um estudo exploratório. *Avaliação (campinas)*, 17(2), 529-556.
- Pfund, R., Norcross, J., Hailstorks, R., Stamm, K., & Christidis, P. (2018). Introduction to psychology: course, purposes, learning outcomes, and assessment practices. *Teaching of Psychology*, 45(3), 213-219. doi:10.1177/0098628318779257
- Price, M., Carroll, J., O'Donovan, B. and Rust, C. (2011). If I was going there I wouldn't start from here: A critical commentary on current assessment practice. *Assessment & Evaluation in Higher Education*, 36(4), 479-92.
- Rawson, K., Dunlosky, J., & Sciartelli, S. (2013). The power of successive relearning: improving performance on course exams and long term retention, *Educational Psychology Review*, 25, 523-548. doi:10-1007/s10648-013-9240-4

- Raymond, J., Homer, C., Smith, R. & Gray, J. (2012). Learning through authentic assessment. An evaluation of a new development in the undergraduate midwifery curriculum. *Nurse Educational and Practice*, 13(5), 471-476. doi:10.1016/j.nepr.2012.10.006
- Roser, M., & Nagdy, M. (2018). *Projections of future education*. Published online at OurWorldInData.org. Available from <https://ourworldindata.org/projections-offuture-education>.
- Saye, J. (2013). Authentic Pedagogy: its presence in social studies classrooms and relationship to student performance on state-mandated tests. *Theory & Research in Social Education*, 41, 89-132. doi:10.1080/00933104.2013.756785
- Struyven, K., & Devesa, J. (2016). Students' perceptions of novel forms of assessment In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 129-144). New York: Routledge.
- Tholen, G., James, S., Warhurst, C., and Commander, L. (2016). Higher education, graduate skills and the skills of graduates: the case of graduates as residential sales estate agents. *British Educational Research Journal*, 42(3), 508-523, doi:10.1002/berj.3222
- Thornburn, M. (2008). Articulating a Merleau-Pontain phenomenology of physical education: The quest for active student engagement and authentic assessment in high-stakes examination awards. *European Physical Education Review*, 4(2), 263-280.
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C. & Herrera-Seda, C. (2018). Authentic assessment: creating a blueprint for course design, *Assessment & Evaluation in Higher Education*, 43(5), 840-854. doi:10.1080/02602938.2017.14123969
- Villarroel, V., Boud, D., Bloxham, S., Bruna, D., & Bruna, C. (2020). Using principles of authentic assessment to redesign written examinations and tests, *Innovations in Education and Teaching International*, 57(1), 38-49, doi: 10.1080/14703297.2018.1564882
- Vu, T. & Dall'Alba, G. (2014). Authentic Assessment for Student Learning: An ontological conceptualization, *Educational Philosophy and Theory*, 46(7), 778-791, doi:10.1080/00131857.2013.795110
- Watkins, D., Dahlin, B., & Ekholm, M. (2005). Awareness of backwash effect of assessment: A phenomenographic study of the views of Hong Kong and Swedish lectures. *Instructional Science*, 33, 283-309. doi:10.1007/s11251-005-3002-8.
- Wiggins, S., Hammar, E., Larsson, G., Pauli, R., & Worrell, M. (2016). Ask Not Only 'What Can Problem-Based Learning Do For Psychology?' But 'What Can Psychology Do For Problem-Based Learning?' A Review of the Relevance of Problem-Based Learning for Psychology Teaching and Research. *Psychology Learning & Teaching*, 15(2), 136-154.
- Zaim, M., Refnaldi, & Arsyad, S. (2020). Authentic Assessment for Speaking Skills: Problem and Solution for English Secondary School Teachers in Indonesia. *International Journal of Instruction*, 13(3), 587-604. doi.10.29333/iji.2020.13340a.