Loughborough University

This item was submitted to Loughborough's Research Repository by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

The importance of controlling item behaviour and scoring methodology for testing higher level learning

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© Loughborough University

VERSION

AM (Accepted Manuscript)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Mackenzie, Don. 2019. “The Importance of Controlling Item Behaviour and Scoring Methodology for Testing Higher Level Learning”. figshare. https://hdl.handle.net/2134/5365.

Loughborough University

This item was submitted to Loughborough's Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

creative commons

COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:



Attribution. You must attribute the work in the manner specified by the author or licensor.



Noncommercial. You may not use this work for commercial purposes.



No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license).

Disclaimer

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

THE IMPORTANCE OF CONTROLLING ITEM BEHAVIOUR AND SCORING METHODOLOGY FOR TESTING HIGHER LEVEL LEARNING

Don Mackenzie

The Importance of Controlling Item Behaviour and Scoring Methodology for Testing Higher Level Learning

Don Mackenzie
Innovation 4 Learning
University of Derby
d.mackenzie@derby.ac.uk

Abstract

When testing higher level learning it is important to ensure that e-assessment is as reliable and as rigorous as possible and that guess-factors are reduced to the minimum in order to be confident that the final score is a true reflection of the candidate's ability.

A number of assessment systems, such as those provided by VLEs, provide only basic controls on question behaviour and scoring together with relatively simple 'click & pick' item types. For more advanced e-assessment it is necessary to employ a dedicated e-assessment system that allows fine control of how each item is delivered and flexible scoring methodologies that allow for the award of partial credit for partially correct answers and stepwise accreditation of the route to the answer even though the final answer may not be correct.

The 'move object' interaction type available in the QuickTri and TRIADS assessment systems illustrates this type of functionality. This interaction type, in its simplest form, may be used to label up a diagram with n labels being moved to n target positions on the diagram. This setup however provides relatively high guess factors and a limited range of scoring nodes in the upper regions of the score range. The guess factor can be reduced by adding dummy labels or even dummy target positions although there are some scoring pitfalls to be aware of here whereby it is possible to score a candidate who makes more errors more highly than one who makes less errors but does not guess. Scoring nodes in the upper regions can be reinstated by providing stacked multiples of each label. The ability to be able to score each label individually in each position together with the appropriate use of negative scoring can enhance the provision of partial credit and a reduction in the guess factor if done carefully.

The rigour of this item type can be enhanced by introducing each of the labels, including any dummy labels, onto the screen one at a time in random sequence. The next label will not appear until the candidate has either moved the previous label into a target or discarded it into a 'waste-bin' target whereupon it will no longer appear on screen. Controlling the question

behaviour in this way makes the question more rigorous because the candidate is prevented from seeing the whole list of labels on screen and less likely to enhance their probability of guessing a proportion of the answer by positioning the labels that they know to be correct then guessing the positions of a small number of the remainder.

The rigour can be further enhanced if the system locks the labels into their target positions so that they cannot be moved once positioned. Such item types can be very useful when testing critical decision-making such as might occur in medical and health and safety assessments. In the full TRIADS Professional system, labels may be presented to the candidate in groups in random sequence for the candidate to position. The next group of labels will not appear until all the labels in the previous group have been positioned or discarded. This question has been used to good effect in the continuing professional development of nurses where candidates are asked to classify drugs with similar sounding names according their very different application. This is a significantly rigorous question when labels are locked into targets so that only their first choices count.

Testing and scoring routes to answers can be undertaken using groups of labels for each step in the route to the ultimate answer with the group for each step appearing after the previous step has been completed and the answer locked into position.

This interaction type can also be used to give an indication of the level of confidence that the candidate has in their answer. By selecting to score the answer history, it is possible to score each positioning of each label cumulatively until the candidate selects to submit their final answer. By careful scoring of each label in each target and intelligent use of internal negative scoring it is possible to identify those candidates who can readily answer the question by correctly positioning all labels first time from those who may also gain the correct answer but who shuffle labels between positions on the way, due to some degree of uncertainty. This of course is a very crude guide to the level of confidence in the answer and a more accurate assessment may require the use of the time taken to answer the question as an additional indicator.

Outside the free-text constructed response interaction, the ‘move object’ interaction type is one of the most flexible currently available for rigorous higher level testing. It can be used for labelling, graphic matching, extended matching item, multiple matching-item and by means of grouping or sequencing targets it is also possible to use it for classification, sequencing and testing sequences within classes.

In formative assessment, uses of this interaction type are legion and provide a very powerful learning tool. Instant, context sensitive feedback on a positioning may be given if it can be configured to be individual to each label in each position. The interaction may be set to ‘scored formative’ mode if the labels are additionally set to be put back if incorrectly positioned and scoring is set to score the history.

However, little of this is possible without appropriate controls on question behaviour and intelligent use of flexible scoring mechanisms underpinned by knowledge of the pitfalls therein.