Chapter 28
Analyzing Process Data from Technology-Rich Tasks

Lisa Keller
University of Massachusetts – Amherst, USA

April L. Zenisky
University of Massachusetts – Amherst, USA

Xi Wang
University of Massachusetts – Amherst, USA

ABSTRACT

A key task emerging in item analysis is identification of what constitutes valid and reliable measurement information, and what data support proposed score interpretations. Measurement information takes on many forms with computerized tests. An enormous amount of data is gathered from technology-based items, tracing every click and movement of the mouse and time stamping actions taken, and the data recorded falls into two general categories: process and outcomes. Outcomes are traditional scored answers that students provides in response to prompts, but technology-based item types also provide information regarding the process that students used to answer items. The first consideration to the practical use of such data is the nature of the data generated when learners complete complex assessment tasks. The chapter we propose serves to discuss some possible methodological strategies that could be used to analyze data from such technology-rich testing tasks.

INTRODUCTION

As assessment increasingly moves toward computerization, technology-rich items are being evaluated to measure academic outcomes that are historically challenging to assess with traditional test formats, such as critical thinking skills. Technology-rich items typically involve one or more of the following elements such as digital media, electronic resources (authoritative literature, websites, etc.), a range of tasks and varied response actions, and high interactivity. Large-scale implementation of these items depends on the capabilities of computer technology to seamlessly deliver such items to examinees given high-levels of administration requirements, and beyond that, a key task

DOI: 10.4018/978-1-4666-9441-5.ch028

emerging in item analysis is the identification of what constitutes valid and reliable measurement information. With technology-rich items, a great deal of data can be generated, but the challenge remains as to what data support proposed score interpretations, and what is, quite simply, noise.

Measurement information can take on many forms with technology-rich items administered via computer. Truly, an enormous amount of data is available for collection with many technology-based item formats, tracing every click and movement of the mouse and time stamping actions taken. The data typically available for recording falls into two general categories: *process* and *outcomes*. Outcomes are traditional scored answers that students provides in response to prompts, but technology-based item types potentially also provide information regarding the *process* that students used to answer items. This process-related information can be used as an element in the scoring of the item, or to provide diagnostic and formative information to educators *and* test developers. The challenge that is emerging with these tasks is distilling the *useful* information from such items; again, separating signal from noise.

Since traditional item types typically do not provide process information, the literature on strategies to analyze and use such data effectively is in its infancy, and thus peer-reviewed guidance is limited. Where such documentation exists is primarily in the form of technical reports advanced by testing agencies specific to particular innovative testing initiatives, and as such is sporadic in nature. Recent work has detailed some of the analyses undertaken in the context of the assessment of critical thinking, while a number of other studies have focused on additional data mining strategies (Almond, Deane, Quinlan, Wagner, & Sydorenko, 2012; Carr, 2013; Bouchet, Harley, Trevors, & Azevedo, 2013; Kinnebrew, Loretz, & Biswas, 2013; Kerr & Chung, 2012). Reflecting on this nascent body of literature, it is possible to identify certain emerging trends relevant to the present work.

The first consideration to the practical use of such data is the nature of the data generated when learners complete complex assessment tasks. At present, the proverbial "anything" is theoretically *possible* but operationally challenging. Technology allows testing programs to gather a vast amount of data during test administration, and this data takes on many forms. It can include records of mouse position, display time of on-screen materials, mouse clicks, drag and drop movements, deletions, drafts, and submitted final products, among other data elements. For operational testing programs, however, collection within this universe of data possibilities presents two issues. First, the data to be gathered must be operationally defined and its purposes articulated. This is necessary because this kind of formal reflection serves as the basis of the technical specification for the data collection mechanism, which must be programmed into the test administration system so as to provide the data miners with the data in a usable (or at least transformable) format. The other issue to be resolved here is that depending on what data is collected and what format it takes, the individual records for any one test session may be quite extensive, to the extent that across examinees the data storage requirements may be substantial and perhaps prohibitive for larger testing programs absent clear goals and plans for the data.

To this end, while process information has potential to be pedagogically quite interesting, techniques to capture and effectively use such data have not been well developed. These data have been characterized as "semi-amorphous" (Scalise, 2013), a term which seems to aptly captures both the challenge and the promise that characterizes innovations in assessment task formats This chapter presents some possible methodological strategies that could be used to analyze data from such technology-rich testing tasks. First, a synthesis and evaluation of the approaches that have been used is presented. The strengths and limitations of the methodologies will be noted, along with the con-