Appendix 2: External Measures to Validate Stealth Assessments

Performance-Based Measure of Persistence

We have developed a performance-based measure of persistence (PBMP) that measures how much effort people exert in difficult tasks (see Ventura, Shute, and Zhao 2012). The PBMP is administered online (in an Internet browser), and presents a variety of hard and easy problems (e.g., anagrams or picture comparison tasks) one at a time over a series of trials. Individuals type in their response and press the "guess" button. If the answer is wrong, the screen displays "incorrect" and the individual can try again (for up to 120 seconds). At any time the individual can also choose to select the "skip" button to leave the current trial and go on to the next one. If the individual guesses correctly, the person is told that he or she is correct. A trial is classified as "solved" if the person accurately completes the trial. A trial is classified as unsolved if the person skips the trial or is timed out after 120 seconds. We propose that persistence may motivate individuals to expend extra effort in solving hard problems outside their ability level. Specifically, the critical information in the PBMP that informs the assessment of persistence is *time spent on unsolved trials*. While the time spent on solved trials is likely a function of persistence as well, it may be dependent on background knowledge or ability in relation to the respective problem. Below are two screen captures: one of a "hard" anagram item (the correct answer is quisby), and one of a "hard" picture comparison task where five differences must be detected between the two pictures—four of which are fairly easy, and one of which is nearly impossible to find (see figure 16).

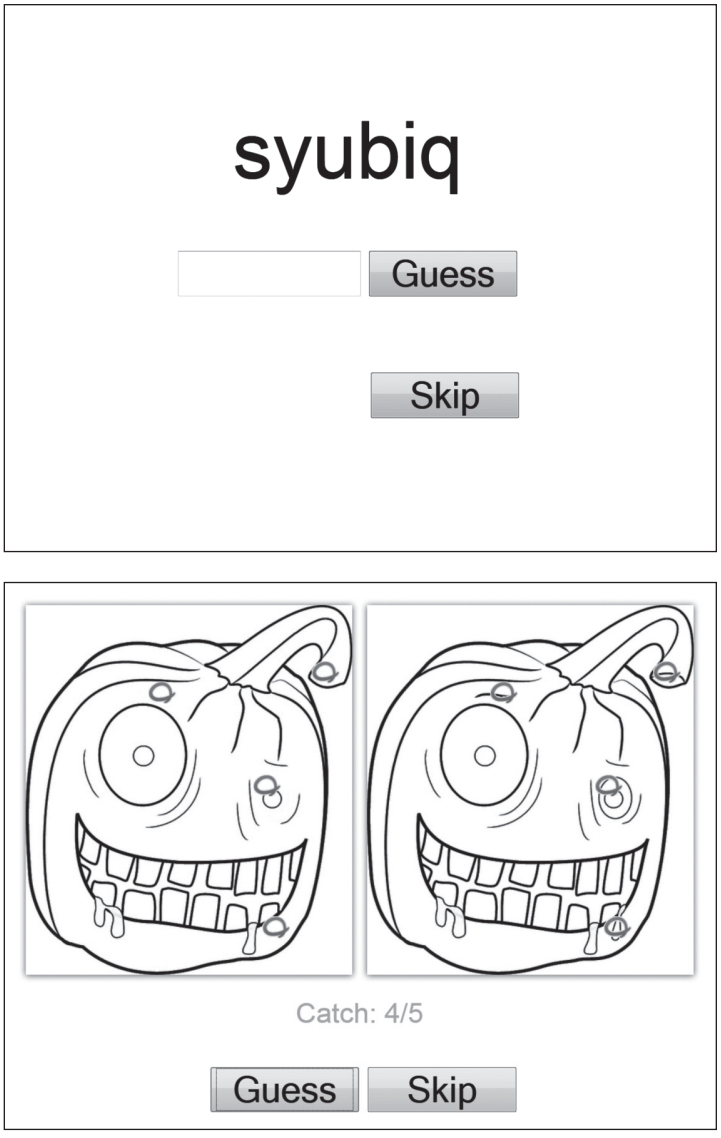

Figure 16

A PBMP has several potential advantages over traditional self-report measures. First, a PBMP can be seen as an implicit measure since no explicit questions are used that may cue the intentions of the assessment. This can mitigate the social desirability effect seen in self-report measures. Second, a PBMP can be claimed to have greater face validity than self-report measures. Self-report measures of persistence ask individuals how they act in difficult problems (e.g., "I never give up"), while performances-based assessments can actually measure behavior in real time on difficult problems. Finally, by nature of their implicit nature and face validity, a PBMP can offer a framework to assess learning of the target construct over time. For example, assessing the effectiveness of an intervention that tries to improve persistence might by compromised by using self-report measures (e.g., social desirability effects, different interpretation of items, lack of explicit knowledge of dispositional change). Alternatively, a PBMP may represent a more appropriate means to evaluate if persistence can be affected due to an experimental manipulation or lifestyle choice (e.g., playing video games).

Self-Report Items

We also plan to use a validated measure from the International Personality Item Pool (IPIP). Students will respond to the following 20 items on a 1–5 Likert scale (1 = strongly disagree to 5 = strong agree; asterisks denote items that are reverse keyed):

Persistence

1. I push myself very hard to succeed
2. I accomplish a lot of work
3. I have patience when it comes to difficult problems
4. I get easily frustrated on new problems*

5. I tend to give up easily*
6. I do more than what's expected of me
7. I tend to avoid difficult problems*
8. I put little time and effort into my work*
9. I enjoy a good challenge
10. I always try my hardest

Perfectionism

1. I dislike routine*
2. I pay attention to details
3. I continue until everything is perfect
4. I have an eye for detail
5. I want every detail taken care of
6. I dislike imperfect work
7. I want everything to add up perfectly
8. I detect mistakes
9. I demand quality
10. I prefer to just let things happen*

Performance-Based Measure of Creativity

Our external measures of creativity will include tests developed and validated by Wallach and Kogan as well as self-report items from the openness survey (also from the IPIP). For the former, we will follow a methodology developed by Wallach and Kogan for their widely used creativity test, which also has good psychometric properties (Wallach 1971). The instrument consists of three verbal tests (instances, alternate uses, and similarities) and two figural tests (abstract patterns and straight or curved lines) concerning the uses of common objects. We will use a version of

their alternate uses test that asks questions such as "Tell me all the different ways you could use a book."

To ensure the reliability and validity of our version of the Wallach and Kogan creativity test, we will follow the scoring framework suggested by Paul Silvia and his colleagues (Silvia et al. 2008). That is, participants will be asked to circle their two most creative responses, and then human raters will judge their responses using a 5-point Likert scale (1 = not at all creative to 5 = highly creative). Based on the ratings provided by human raters, two creativity indexes will be used for the overall creativity score: the average creativity index (i.e., the sum of ratings across all responses divided by the number of responses), and the rating for the top two responses.

Self-Report Items

We plan to use a validated measure from the IPIP. Students will respond to the following 10 items on a 1–5 Likert scale (1 = strongly disagree to 5 = strong agree):

Openness

1. I like to think of new ideas
2. I enjoy art
3. I am excited by many different activities
4. I daydream a lot
5. I enjoy learning new things
6. I like to explore different solutions to problems
7. I have an active imagination
8. I like to be original
9. I try to be different from other students
10. I am curious about many different things

External Assessment of Conceptual Physics

We currently have a set of twenty-four items (i.e., twelve items in form A, and twelve isomorphic items in form B) that assess the competencies in our conceptual physics competency model. The test is divided among the four main agents of force and motion. Within each section, different facets of the physics principles are assessed. Items are either multiple choice or constructed response. For constructed response items (like the one shown in figure 17), our rubric will consist of an optimal trajectory surrounded by an area comprising a "correct response."

For our multiple-choice items, the format is generally the same—where a problem is presented, along with a graphic that

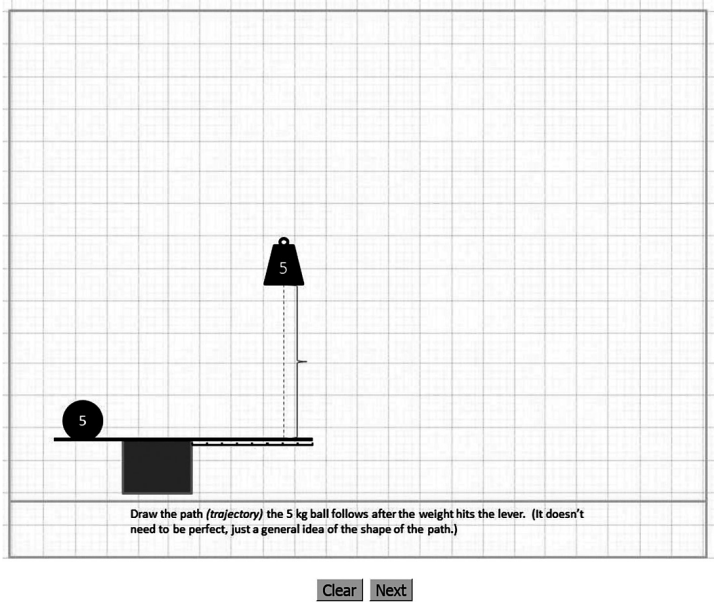

Figure 17

consists of two options (A and B). The student has to decide whether the graphic depicted in A is correct, B is correct, both A and B are equal, or the answer isn't known. See the example shown in figure 18.

The items were created based on Hewitt's 2009 textbook (*Conceptual Physics*, eleventh ed.), and then reviewed and edited by our physicist working on the project (Dr. Donald Franceschetti, University of Memphis). We also plan to use items from the Force Concept Inventory (Hestenes, Wells, and Swackhamer 1992) to measure for the transfer of physics principles. For example, we expect that playing *Newton's Playground* will result in the conceptual understanding of various object collisions (e.g., moment of inertia) not explicitly observed in *Newton's Playground*.

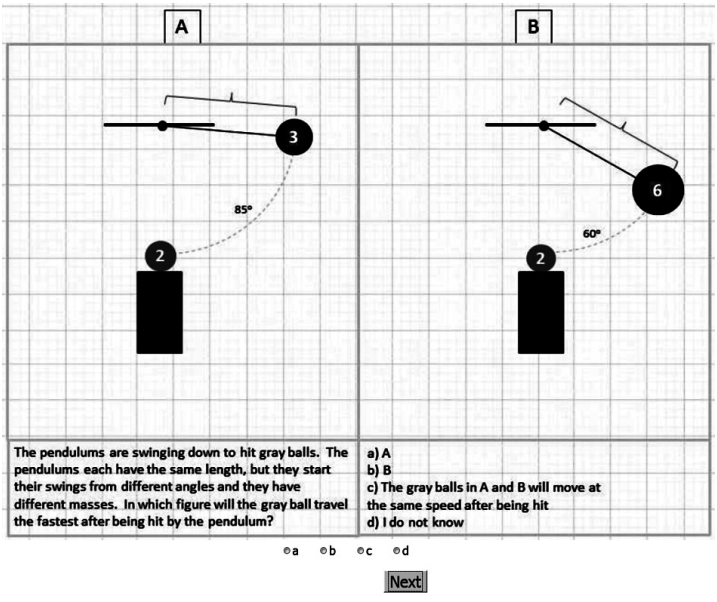

Figure 18

This is a section of doi:10.7551/mitpress/9589.001.0001

Stealth Assessment

Measuring and Supporting Learning in Video Games

By: Valerie Shute, Matthew Ventura

Citation:
Stealth Assessment: Measuring and Supporting Learning in Video Games
By: Valerie Shute, Matthew Ventura
DOI: 10.7551/mitpress/9589.001.0001
ISBN (electronic): 9780262315210
Publisher: The MIT Press
Published: 2013

The MIT Press

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Stone Serif and Stone Sans by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data
Shute, Valerie J. (Valerie Jean), 1953– , author.
Stealth assessment : measuring and supporting learning in video games / Valerie Shute and Matthew Ventura.
pages cm. — (The John D. and Catherine T. MacArthur Foundation reports on digital media and learning)
Includes bibliographical references.
ISBN 978-0-262-51881-9 (pbk. : alk. paper)
1. Educational tests and measurements. 2. Video games. I. Ventura, Matthew, author. II. Title.
LB3051.S518 2013
371.26—dc23
2012038217

10 9 8 7 6 5 4 3 2 1