Assessing Computer Literacy of Adults with Low Literacy Skills

Andrew M. Olney
Institute for Intelligent Systems
University of Memphis
Memphis, TN 38152
aolney@memphis.edu

Dariush Bakhtiari
Department of Educational Psychology, Special Education, and Communication Disorders
Georgia State University
Atlanta, GA 30302
dbakhtiari1@gsu.edu

Daphne Greenberg
Department of Educational Psychology, Special Education, and Communication Disorders
Georgia State University
Atlanta, GA 30302
dgreenberg@gsu.edu

Art Graesser
Institute for Intelligent Systems
University of Memphis
Memphis, TN 38152
a-graesser@memphis.edu

ABSTRACT
Adaptive learning technologies hold great promise for improving the reading skills of adults with low literacy, but adults with low literacy skills typically have low computer literacy skills. In order to determine whether adults with low literacy skills would be able to use an intelligent tutoring system for reading comprehension, we adapted a 44 task computer literacy assessment and delivered it to 114 adults with reading skills between 3rd and 8th grade levels. This paper presents four analyses on these data. First, we report the pass/fail data natively exported by the assessment for particular computer-based tasks. Second, we undertook a GOMS analysis of each computer-based task, to predict the task completion time for a skilled user, and found that it negatively correlated with proportion correct for each item, $r(42) = -.4,\ p = .01$. Third, we used the GOMS task decomposition to develop a Q-matrix of component computer skills for each task, and using logistic mixed effects models on this matrix identified five component skills highly predictive of the success or failure of an individual on a computer task: function keys, typing, using icons, right clicking, and mouse dragging. And finally, we assessed the predictive value of all component skills using logistic lasso.

Keywords
adult literacy, computer literacy, GOMS, Q-matrix, mixed model, lasso

1. INTRODUCTION
Of adults with the lowest literacy levels, 43% live in poverty, and low literacy costs the U.S. economy $225 billion annually [14]. The need for literacy interventions is matched by the complexity of delivering interventions to this population. Low literacy adults have difficulty attending face to face programs at literacy centers because of work, child care, and transportation [5], and even when these challenges are met, two-thirds of literacy centers have long waiting lists [14]. Adaptive computer-based interventions for literacy hold promise to overcome these challenges. Such interventions can be deployed in homes and local libraries, in addition to literacy centers. However, computer-based interventions raise another question: can adults with low literacy skills use computers well enough to benefit? Several surveys suggest that this might be a problem. The demographics most affected by low literacy are the same demographics least likely to use the Internet (over age 50, making less than $30 thousand a year, and with less than a high school education [1]).

Several decades of research have investigated computer literacy using self-report measures as well as objective tests, i.e. multiple choice, and find that self-report measures tend to exaggerate proficiency while objective tests are more reliable (see [3] for a review). For an adult literacy population, however, multiple-choice tests delivered as print create additional concerns as to whether the questions themselves can be comprehended. Recently a new type of assessment, known as the Northstar Digital Literacy Assessment (the Northstar), has been created that directly measures ability to perform computer tasks [13]. Unlike multiple choice assessments, the Northstar can simulate a computer desktop, use voice prompts to instruct users to perform tasks on that desktop, and then record their mouse clicks and keystrokes to determine if the task has been completed. Almost all of the tasks can be completed without reading by listening to the voice prompt instructions. The few tasks that do involve reading are word recognition tasks rather than sentence reading, e.g. a task to log in may require the user to copy a name and password to the appropriate boxes and so require reading of “Username,” “Password,” and the corresponding fillers. The Northstar has been adopted as the computer literacy standard for adult basic education in the state of Minnesota, which further supports its appropriateness for assessing the computer literacy skills of adults with low literacy skills.

The present study investigated the computer literacy skills of adults with low literacy skills for the purpose of developing an intelligent tutoring system for reading comprehension for this population [7]. It includes a set of Northstar items that were collected to cover a range of potential interface and interaction components. In the remainder of the paper we describe the data collection procedure and four analyses performed, including pass/fail frequencies for each task, relation of these frequencies to GOMS-predicted execution times for skilled users, a logistic mixed-model using a Q-matrix decomposition of the tasks into component skills, and a logistic lasso model to assess the predictive value of component skills. From these analyses we identify specific tasks that are problematic for adults with low literacy skills as well as component skills that make it more likely adults with low literacy skills will succeed or fail at a computer-based task.

2. ANALYSIS 1: PROPORTION CORRECT

2.1 Participants

Participants (N = 114) were recruited through adult literacy centers in Atlanta, GA and Toronto, ON, from classes where the reading level was between 3rd and 8th grade. Reading level was determined by the centers using their "business as usual" assessments. Demographic surveys were completed by 90 participants (79% completion rate). Completed surveys indicated that participants were slightly more female than male (55 vs. 35) and that participant age ranged from 17 to 69 ($M = 42.74,\ SD = 13.73$).

2.2 Materials

Forty-four items were selected from four (out of seven) of the Northstar modules available at the time of the study: Basic Computer Skills (21), WWW (13), Windows (6), and Email (4). Task descriptions are given in Table 1. Basic Computer Skills covered such topics as turning a computer on, identifying components of a computer, files and folders, menus, and windows. WWW focused on browser-based activities like searching, search results, browser functionalities, and logging in. Although the Windows module focused on Windows overall, the items selected were fairly generic to any windowed operating system and mostly pertained to desktop applications. Email questions used a webmail interface (browser-based email client) and queried how one would create a new email, send an email, or similar email task. Because Northstar modules are integrated assessments, the Northstar Project compiled the items we selected into a custom assessment for us.

2.3 Procedure

Participants first completed informed consent and then the demographic survey. Both informed consent and demographic survey were read aloud to participants to ensure comprehension. Participants were then asked to sit in front of a computer to take the Northstar assessment. The assessment was delivered in the browser using Adobe Flash. At the start of the assessment, a 3-minute orientation video was played explaining how to answer questions in the assessment. If the participant was confused about what to do, an experimenter was available to answer questions. Each question consisted of an voice prompt defining the task, which was also written at the top of the screen. A replay button was available to repeat the prompt. Participants could select, click, type, drag, etc. on the interface in an attempt to perform the task. If the participant did not know how to complete the task, they could press an "I Don't Know" button, at which point the system scored their attempt as a failure. Attempts were only scored as a success if the participant completed the task in the manner requested in the prompt. The completion of each task initiated the next task until the assessment was complete.

2.4 Results & Discussion

The Northstar records success/failure of each participant on each task, and these data are reported in detail elsewhere [2]. Here we briefly note that the proportion of correct responses for each task is quite wide, ranging from .19 to .98. Tasks in which participants performed particularly well (proportion correct above .80) include identification tasks (e.g. for mouse, keyboard, headphone jack, and websites), turning on a computer or monitor, and common operations like recycling a file, using checkboxes, dragging, scrolling, and using hyperlinks. Tasks in which participants performed poorly (proportion correct below .60) include identification of various keys, double- or right-clicking, typing web addresses, signing into email, and composing email.

The proportion correct results from the Northstar indicate the adults with low literacy skills can power on their device and perform a variety of basic operations. To the extent that these tasks exactly matched tasks that would be performed in a computer-based literacy intervention, like an intelligent tutoring system, this level of results is quite useful. However, for some tasks there is not an exact match, and the implications of the proportion correct results are less clear. For example, difficulties performing tasks using Word, Excel, or webmail may reflect problems with those specific interfaces that may not transfer to other programs. Understanding these more nuanced relationships would require a deeper analysis than is afforded by Northstar's success/failure output.

3. ANALYSIS 2: GOMS MODELING

The purpose of this analysis was to explore whether the success rate of the Northstar tasks could be modeled using GOMS (Goals, Operators, Methods, & Selection rules), a well-known computational technique for modeling expert user performance on a task [10]. GOMS decomposes a particular computer task, e.g. saving a file, into goals and subgoals, perceptual, cognitive, and motor actions in service these goals, methods or sequences of operators that achieve a goal, and selection rules that choose between alternative methods. An important assumption of GOMS is that the users are expert at the computer task in question. Therefore GOMS models of execution time represent the upper bound of performance after a user has learned the interface and practiced it many times. The expert assumption of GOMS is violated in the adult literacy population, making the outcome of this analysis non-obvious. If the GOMS model predictions of execution time were related to our adult's performance, that would provide evidence that GOMS modeling

Table 1: Northstar Tasks

Click on the monitor	Recycle file	Click stop loading
Click on the keyboard	Checkboxes	Select search engines
Click on the system unit	Organize folder options	Google query
Click on the headphone jack	Start menu, lauch program	Google scroll
Click on picture of a mouse	Turn up audio slider	Use hyperlink
Newline key	Mute audio	Maximize window
Caps key	Select browser icons	Minimize window
Shift key	Click on the website	Open Excel
Backspace key	Drag item in browser	Open Word using taskbar
Up arrow	Click on address bar	Close Word
Turn on monitor	Type the web address	Select login and password
Turn on computer	Click homepage button	Choose secure password
Log on to computer	Click browser back button	Sign into email
Double click on Documents	Click browser refresh	Compose email
Right click menu	Click browser forward	

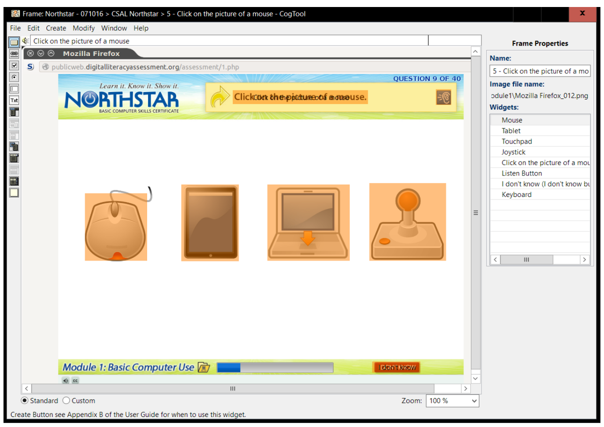

Figure 1: A CogTool annotation of a Northstar task. Annotations appear as semi-transparent orange boxes over the Northstar interface.

has some validity for this population.

3.1 Procedure

The CogTool system was used to perform a GOMS analysis [11, 9]. CogTool allows the easy creation of GOMS models by annotating an existing user interface, and then recording a demonstration of the task against than annotated interface. Figure 1 shows the CogTool interface for the "Click on the mouse" task. For example, when the Northstar task required clicking on an icon, button, or other interface element as in Figure 1, a CogTool button annotation was overlaid on the interface, and then in demonstration mode the modeler would demonstrate the task by clicking on the annotated button. From this demonstration on the annotation, CogTool builds a GOMS model that includes the perceptual, cognitive, and motor tasks required to perform the task. Similar annotations were made for auditory directions, keyboard input, and other kinds of interface actions. Once a task was annotated and demonstrated, a CogTool simulation was run on GOMS model to generate a predicted execution time of expert performance. Annotations, demonstrations, and execution time predictions were performed for all 44 Northstar items used in Analysis 1.

3.2 Results & Discussion

GOMS-predicted execution times for Northstar tasks ranged from 3.0 to 17.1 seconds ($M = 6.88,\ SD = 4.07$). These execution times were significantly negatively correlated with proportion correct, $r(42) = -.40,\ p = .01,\ CI_{95}[-.61, -.10]$, indicating that tasks predicted to take an expert longer to accomplish were more likely to be answered incorrectly by low literacy adults. Tasks that take longer are inherently more complex and require more operations to complete. These results suggest that GOMS has some validity for modeling the performance of adults with low literacy skills even though it was not intended for this purpose. However, by themselves these results convey little additional insight. The GOMS-predicted execution times, generated by CogTool, are still at the task level rather than the component skills required to achieve each task. This is partly because the orientation of CogTool is to produce execution times and partly because of the expert orientation of GOMS. For example, in GOMS the factors involved in clicking a button are the perceptual (size, location) and motor operations involved, but in Northstar, some "buttons" are tapping specific types of knowledge, like identifying hardware, understanding icons, or various keys on a keyboard. The different types of knowledge behind the various CogTool annotations are not represented or considered in the GOMS analysis it provides.

4. ANALYSIS 3: Q-MATRIX & LOGISTIC MIXED MODELS

We would like to understand how the component skills underlying Northstar tasks differentially affect the probability a low literacy adult will perform the task correctly. In educational data mining, component skills are typically modeled using a Q-matrix analysis [4]. In its simplest form, a Q-matrix analysis constructs a problem by skill matrix such that a $cell_{ij}$ in the matrix represents whether $skill_i$ is needed to solve $problem_j$: $cell_{ij} = 1$ if $skill_i$ is needed to solve $problem_j$, and $cell_{ij} = 0$ if $skill_i$ is not needed to solve $problem_j$. Analysis 2 provides a useful guide towards the creation of a Q-matrix for the Northstar tasks, as it has already captured each component action required to perform

Table 2: Component skills coded from GOMS

Component Skill	Probability Correct Given Skill
Checkboxes	.89
Mouse Drag	.86
Hardware Identify	.83
Hardware Function	.78
Complex Scrolling	.74
Browser Functions	.66
Left Click	.64
Use Icons	.61
Double Click	.58
Window Functionality	.56
Program Brands	.55
Desktop Concept	.53
Select Menu	.50
Good Login Info	.50
Login Info	.48
Keyboard Function	.46
Simple Typing	.43
Right Click	.19

each task. What it lacks in some cases, however, is an annotation of the knowledge behind each component action.

4.1 Procedure

The first author recoded the GOMS task annotations with 18 novice-relevant component skills. The coding was done in one pass, and component skills were defined on the fly. Component skills that occurred in only one task were then removed as they offer no predictive utility for other tasks. The appropriateness of the component skills was evaluated by correlating the total number of component skills needed in each task with the GOMS execution time and the proportion correct for the respective task. We used a logistic mixed model to predict the correctness of each participant on each task as a function of the presence of component skills for that task. This analysis addresses the question as to whether there is an effect (main effect) of the presence of component skills on the likelihood that an adult with low literacy skills will be able to perform the task correctly. Using a logistic mixed model in this way has strong similarities to cognitive psychometric models like Diagnostic Classification Models [16] or more specifically a mixed model implementation of linear logistic test models [15].

In the logistic mixed model, random slopes were initially included but failed to converge. Random intercepts for task and participant are theoretically motivated, and backward selection of these effects using Akaike information criterion (AIC) achieved a minimum when these effects were included, indicating that these intercepts should remain in the model. These random intercepts can be considered as per-task difficulty not captured by component skills and per-subject ability, respectively. The initial model that included Left Click was rank deficient, so Left Click, which appears in most tasks, was removed from the final model. Additionally, the total number of component skills in each task (i.e. column sums of the Q-matrix) was initially considered as a predictor of correctness, but was excluded based on extremely high collinearity, having a variance inflation factor of over 40.

4.2 Results & Discussion

The component skills and the conditional probability that a task will be correctly performed if the component skill is present are shown in Figure 2. Total component skills per task was marginally positive correlated with GOMS execution time, $r(42) = .27$, $p = .07$, $CI_{95}[-.02, .53]$, suggesting that tasks with more component skills take longer to perform. Total component skills per task was significantly negatively correlated with proportion correct, $r(42) = -.35$, $p = .02$, $CI_{95}[-.59, -.06]$, indicating that tasks with more component skills are more difficult to perform correctly. The correlation between predicted execution time and proportion correct was not significantly different from the correlation between total component skills and proportion correct, $t(82) = .18$, $p = .86$, indicating that the Q-matrix decomposition of component skills is comparable to the GOMS execution time in terms of its relationship to proportion correctness. Altogether these correlation results provide additional evidence that the Q-matrix decomposition is appropriate.

The logistic mixed model had a marginal R^2 of .18 (fixed effects only) and a conditional R^2 of .47 (including random effects) [12]. We found a positive main effect of Mouse Drag, $\hat{\beta} = 2.06$, $SE = .90$, $p = .02$, such that tasks with a Mouse Drag component were 7.87 times as likely to be answered correctly, $CI_{95}[1.36, 45.50]$, and a marginal main effect of Hardware Identify, $\hat{\beta} = .89$, $SE = .53$, $p = .10$, such that tasks with a Hardware Identify component were 2.44 times as likely to be answered correctly, $CI_{95}[.86, 6.94]$. We found negative main effects for Keyboard Function, $\hat{\beta} = -1.31$, $SE = .51$, $p = .01$, Use Icon $\hat{\beta} = -1.35$, $SE = .55$, $p = .01$, Simple Typing $\hat{\beta} = -1.91$, $SE = .64$, $p = .003$, and Right Click $\hat{\beta} = -3.20$, $SE = 1.34$, $p < .02$, such that tasks with a Keyboard Function component were .27 times as likely to be answered correctly, $CI_{95}[.10, .73]$, tasks with a Use Icon component were .26 times as likely to be answered correctly, $CI_{95}[.09, .75]$, tasks with a Simple Typing component were .15 times as likely to be answered correctly, $CI_{95}[.04, .52]$, and tasks with a Right Click component were .04 times as likely to be answered correctly, $CI_{95}[.00, .56]$.

We found that Mouse Drag was extremely predictive of success. The reason is unclear, but we hypothesize that the frequency of mouse dragging in many computer tasks may have afforded participants the opportunity to become expert in this skill. Mouse dragging has some similarity to swiping on a smartphone or tablet interface, so it may be that expertise with other devices has transferred into the Northstar tasks. Amongst the components that predict failure, perhaps the most intuitive are Keyboard Function and Simple Typing. Typing is a complex skill that takes practice to master. Function keys are difficult in that they don't themselves produce a character, but either operate on a character on the screen (Delete) or work in combination with another key to modify it (Shift). The negative effects associated with Use Icon and Right Click are somewhat surprising. Icons come in many different variations, and so it is possible that the negative Use Icon effect is attributable to a lack of knowledge of specific icons or perhaps to the conventions of icons generally. Right Click is possibly rare and usually brings up a context menu with commands that are often available elsewhere, making it more relevant for power users but perhaps less so to novice users.

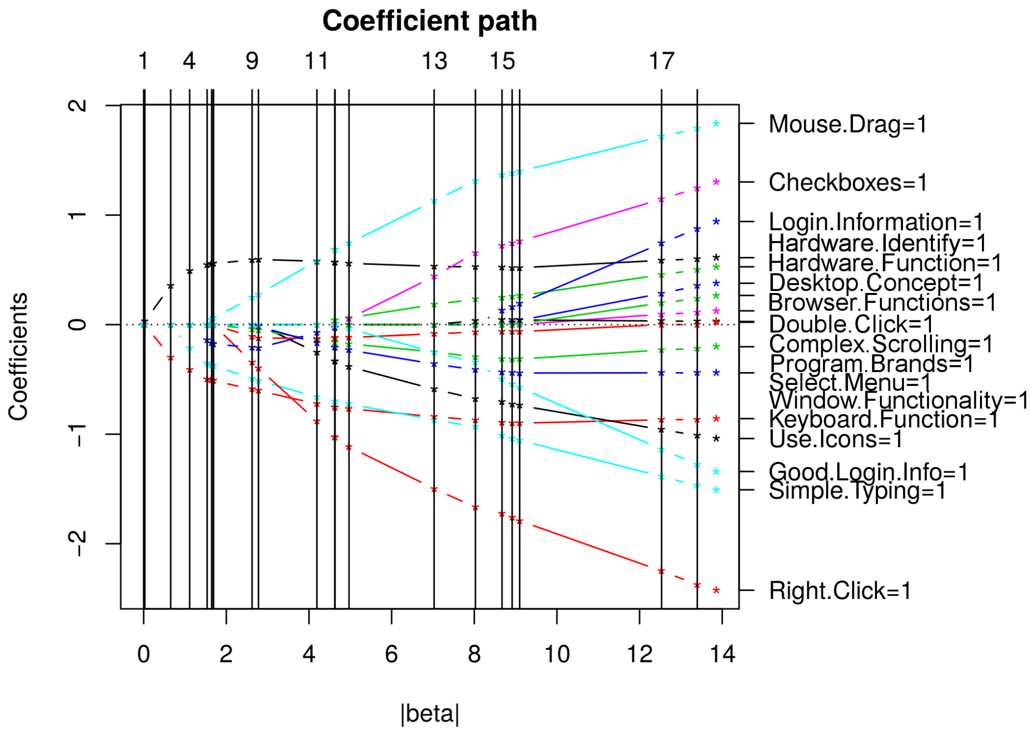

Figure 2: The coefficient path for the lasso model. As the L1 sparsity threshold increases along the x-axis, more coefficients are non-zero.

5. ANALYSIS 4: Q-MATRIX LASSO

Analysis 3 provides a more traditional analysis of significant predictors in our study, but must be interpreted with caution with respect to generalizing to new data. It may be that insignificant predictors in Analysis 3 nevertheless have predictive value on new data. The problems of relying on p-values or criteria like AIC to select variables are well known [8]. To explore the predictive potential of the Q-matrix component skills, we created a lasso model (least absolute shrinkage and selection operator [18]), a form of regression that promotes sparsity (i.e. zero coefficients) and predictive accuracy simultaneously. While not necessarily the best predictive model (cf. gradient boosting [6]), lasso has the advantage of being simple to interpret, and thus our results can guide what variables to use in future models.

5.1 Procedure

A logistic regression base model without random effects was initialized with 17 component skills (Left Click excluded) and submitted to lasso. Because lasso has a free parameter, λ, that controls sparsity of the regression, a lasso analysis varies the level of λ and generates regression coefficient estimates at each level. This sequence of regression coefficients is known as the regularization path. The value of λ that minimized prediction error was estimated using both cross validation and AIC.

5.2 Results & Discussion

The coefficient (regularization) path for the lasso model is shown in Figure 2 and the corresponding AIC curve is shown in Figure 3. In Figure 2, the center line represents coefficients having zero values. As the L1 sparsity threshold ($|beta|$) increases, more coefficients become non-zero. For selecting the optimal λ that minimizes overall prediction error, ten-fold cross validation and AIC yielded congruent results. AIC results are depicted in the curve in Figure 3, which shows that that AIC improves as $|beta|$ increases, coming to a minimum at $|beta| = 13.40$. Accordingly, most coefficients for the optimal lasso model are non-zero.

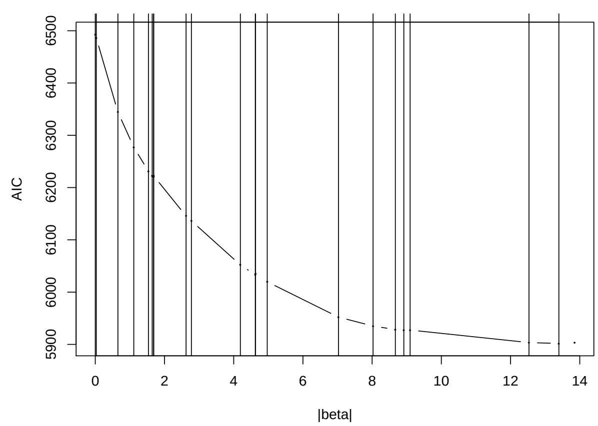

Figure 3: The AIC curve for the lasso model. Lower values of AIC indicate better model fit.

Table 3: Lasso component skill coefficients

Component Skill	$\hat{\beta}$	$exp(\hat{\beta})$
Mouse Drag	1.80	6.02
Checkboxes	1.27	3.55
Login Information	.88	2.41
Hardware Identify	.60	1.82
Hardware Function	.50	1.65
Desktop Concept	.35	1.43
Browser Functions	.24	1.27
Double Click	.11	1.12
Complex Scrolling	.03	1.03
Program Brands	.00	1.00
Select Menu	-.21	.81
Window Functionality	-.44	.64
Keyboard Function	-.86	.42
Use Icons	-1.01	.36
Good Login Info	-1.28	.28
Simple Typing	-1.47	.23
Right Click	-2.39	.09

Table 3 gives the $\hat{\beta}$ coefficients (log odds) for the AIC-optimal model as well as the odds ratio $exp(\hat{\beta})$ for each coefficient. The coefficients converted to odds ratios have the same interpretation as in the logistic mixed model, e.g. tasks with a Mouse Drag component are 6.02 times as likely to be answered correctly as those without. Although the logistic lasso model does not include random intercepts corresponding to task difficulty and subject ability, the magnitudes of coefficients in the logistic lasso are highly comparable to the logistic mixed model. However, the strength of the coefficients in the logistic lasso are weaker, in general, than in the logistic mixed model, suggesting that the logistic mixed model may be slightly over-fitted. For example, according to the logistic mixed model, Mouse Drag tasks are 7.87 times as likely to be answered correctly, but according to the logistic lasso model, Mouse Drag tasks are only 6.02 times as likely to be answered correctly; similarly Right Click containing tasks in the mixed model are .04 times as likely to be answered correctly compared to .09 times as likely in the logistic lasso. These results suggest that while the logistic mixed model might be more appropriate for assessment purposes, as it additionally estimates task difficulty and subject ability, the logistic lasso model might be more appropriate for predicting the effects of component skills on success rates for new tasks.

6. GENERAL DISCUSSION

Together, our results suggest that not only are there specific Northstar tasks that are informative with regard to building an adaptive computer-based intervention for adults with low literacy skills but also that these tasks can themselves be decomposed into component skills that can be further used for this purpose. The main effects of Analysis 3 and coefficient rankings of Analysis 4 are consistent and complimentary with the proportion correct results in Analysis 1. The marginal main effect for Hardware Identify explains the high proportion correctness for identification tasks for mouse, keyboard, and headphone jack, and the main effect for Mouse Drag explains the high proportion correctness for recycling a file (dragging to the Recycle Bin), dragging, and scrolling (by dragging a scroll bar). These correctness-enhancing main effects are also reflected in odds ratios greater than one in Analysis 4. Similarly the main effects for Keyboard Function and Simple Typing explain the low proportion correctness for identifying various keys, typing web addresses, signing into email, and composing email, and these main effects are likewise reflected in odds ratios less than one in Analysis 4. In these cases we infer that the problem is not specific to the interface in question, e.g. email, but rather that there is a deficiency in a component skill needed for the task taking place in the context of that interface.

The implications for building adaptive computer-based interventions for adults with low literacy skills are clear. First, it is important to keep typing to a minimum, either by having users select response options or by using speech recognition. Second, right clicking should be eliminated or at least made optional. Third, icons should be close to icon archetypes. And finally, mouse dragging is a good skill around which to build user interaction. Interestingly, all of these implications seem to point to tablet and smartphone platforms, which have a minimum of typing (and built in speech interfaces), no right clicking, minimal icons in-app, and plenty of swiping/dragging. Moreover, smartphone ownership has been rapidly increasing – now 64% of households earning below $30 thousand own a smartphone [17]. It may be the case that deploying interventions on smartphones and tablets better makes use of both the computer literacy strengths and the material resources of low literacy adults.

7. ACKNOWLEDGMENTS

This research was supported by the Institute of Education Sciences (IES; R305C120001). Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the author and do not represent the views of the IES.

8. REFERENCES

[1] M. Anderson and A. Perrin. 13% of Americans don't use the internet. Who are they? Technical report, Pew Research Center, 2016.

[2] D. Bakhtiari, A. Olney, and D. Greeberg. Computer literacy skills of adult learners. In preparation.

[3] J. A. Ballantine, P. M. Larres, and P. Oyelere. Computer usage and the validity of self-assessed computer competence among first-year business students. *Computers & Education*, 49(4):976 – 990, 2007.

[4] T. Barnes, D. Bitzer, and M. Vouk. Experimental analysis of the q-matrix method in knowledge discovery. In *International Symposium on Methodologies for Intelligent Systems*, pages 603–611. Springer, 2005.

[5] H. Beder and P. Medina. Classroom dynamics in adult literacy education. ncsall research brief. Technical report, National Center for the Study of Adult Learning and Literacy, 2002.

[6] J. H. Friedman. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4):367–378, 2002.

[7] A. C. Graesser, Z. Cai, W. O. Baer, A. M. Olney,

X. Hu, M. Reed, and D. Greenberg. Reading comprehension lessons in AutoTutor for the Center for the Study of Adult Literacy. In S. A. Crossley and D. S. McNamara, editors, *Adaptive Educational Technologies for Literacy Instruction.*, pages 288–293. Routledge, 2016. DOI: 10.4324/9781315647500 DOI: 10.4324/9781315647500.

[8] F. Harrell. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Graduate Texts in Mathematics. Springer, 2001.

[9] B. E. John. Cogtool: Predictive human performance modeling by demonstration. In *Proceedings of the 19th Conference on Behaviour Representation in Modeling and Simulation*, pages 83–84, 2010.

[10] B. E. John and D. E. Kieras. The goms family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3(4):320–351, 1996.

[11] B. E. John, K. Prevas, D. D. Salvucci, and K. Koedinger. Predictive human performance modeling made easy. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 455–462, New York, NY, USA, 2004. ACM.

[12] S. Nakagawa and H. Schielzeth. A general and simple method for obtaining r2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2):133–142, 2013.

[13] N. D. L. Project, 2016.

[14] ProLiteracy. U.S. adult literacy facts. Technical report, 2017.

[15] F. Rijmen, P. D. Boeck, and K. U. Leuven. The random weights linear logistic test model. *Applied Psychological Measurement*, 26(3):271–285, 2002.

[16] A. A. Rupp and J. L. Templin. Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, 6(4):219–262, 2008.

[17] A. Smith. Record shares of americans now own smartphones, have home broadband. Technical report, Pew Research Center, 2017.

[18] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.