Would you? Could you? On a tablet? Analytics of Children's eBook Reading*

Beata Beigman Klebanov
Educational Testing Service
bbeigmanklebanov@ets.org

Anastassia Loukina
Educational Testing Service
aloukina@ets.org

Nitin Madnani
Educational Testing Service
nmadnani@ets.org

John Sabatini
Educational Testing Service
jsabatini@ets.org

Jennifer Lentini
Educational Testing Service
jlentini@ets.org

ABSTRACT

It is difficult to overstate the importance of literacy for adequate functioning in society, from educational attainment and employment opportunities to health outcomes. We created a reading app with the goal of helping readers improve their reading skill while reading for meaning and pleasure, and used it to collect unique data on children's extended reading. Analysis of the data reveals the importance of a behavioral factor in understanding observed reading performance.

CCS CONCEPTS

• **Applied computing** → **Education**; **Interactive learning environments**; • **Information systems** → **Data analytics**; • **Human-centered computing** → Empirical studies in HCI;

KEYWORDS

children's reading ; fluency ; comprehension ; reading app; book reading ; reading analytics ; motivation

ACM Reference Format:
Beata Beigman Klebanov, Anastassia Loukina, Nitin Madnani, John Sabatini, and Jennifer Lentini. 2019. Would you? Could you? On a tablet? Analytics of Children's eBook Reading. In *The 9th International Learning Analytics Knowledge Conference (LAK19), March 4–8, 2019, Tempe, AZ, USA.* Tempe, AZ, USA, Article 000, 5 pages. https://doi.org/10.1145/3303772.3303833

1 INTRODUCTION

Learning analytics, as a field, is seeking insight into learning processes and outcomes by means of collection and analysis of relevant large scale data, which is greatly facilitated by today's digital environments. Indeed, according to [10], 63% of K-12 teachers in the U.S. make daily use of technology in their classroom. This is a golden opportunity to track the processes of learning that are happening over the technological platform (a) unobtrusively, and (b) in rich and minute detail, and seek insight into cognitive, behavioral, social and other aspects of the necessarily complex process of human learning and development.

We describe a study in which we capitalized on exactly such an affordance and were able to track a group of upper elementary school children engaging in a daily independent book reading activity over the course of a summer camp. Analyzing the collected data, we first sought to confirm findings in the reading literature about inter-connection among various reading subskills, by specifically considering the relationship between reading comprehension and reading rate. We found that the relationship is not straightforward, and is mediated by a *behavioral* (that is, not skill based) aspect of the child's long-term engagement with the reading app.

2 DATA COLLECTION INSTRUMENT

According to the recent report from the National Assessment of Educational Progress, thirty-two percent of U.S. 4th graders read below the Basic level [9]. The goal of the project is to help low proficiency readers improve their reading skill through sustained reading with technology-based support, specifically targeting the transition from word-by-word reading to reading fluency. The method is to try and engage the user in the flow and process of reading for meaning and pleasure by (a) choosing a high-interest, fairly long and fairly challenging book, and (b) using technology to enhance engagement, alleviate frustration, and provide feedback. The main idea is to allow the user to take turns reading aloud with a virtual partner, realized, in our case, through an audiobook narration. Our premises are that the interest in the story and the quality of the narration would increase enjoyment, while the interleaving of the effortful reading with the more relaxing experience of listening to a skilled narrator should help reduce the perceived difficulty of the task. Furthermore, due to the continuity of characters, settings, and events throughout the story, many of the actual words the reader would have to read during his or her turn would have been just read by the narrator in his turn, effectively modeling the target reading behavior for the user while also advancing in the story. The web and mobile app MyTurnToRead that implements this idea is described next.

MyTurnToRead is an app that is available both via the web and also as standalone iOS and Android apps (currently in beta). The mobile apps are built using Apache Cordova[1] – a cross-platform toolkit – with platform-specific modifications where necessary. The reading and listening components in the web-app and the mobile

*Paraphrase of *Would you? Could you? In a car?*, "Green Eggs and Ham", by Dr. Seuss.

LAK19, March 4–8, 2019, Tempe, AZ, USA

ACM ISBN 978-1-4503-6256-6/19/03…$15.00
https://doi.org/10.1145/3303772.3303833

[1]https://cordova.apache.org

apps are built on top of Readium,[2] a robust, standards-compliant, and open-source EPUB-based reading system; see Figure 1.

In order to check that the readers are paying attention to the story they are reading and not only "calling words" [7], as well as to remind the reader about the important points in the story, we created more than 700 reading comprehension questions, approximately one question for every 100 words of running text. These are surface-level questions focused on the plot, on relationships between characters, on descriptive detail of characters or events that are important for the story overall; the answers to the questions are generally stated in the text. For the data collection described in this paper, the schedule of reading comprehension questions was as follows: the user was asked two questions after every other user turn, namely, the recurrent unit of reading activity was narrator-user-narrator-user-question-question. Each question is anchored to the string in the text that contained the latest piece of information required to answer the question. The reader is asked the two questions that most closely precede the reader's current bookmark.

As the children are reading with MyTurnToRead, the app logs information about their interaction with the app. The audio of the user's turn is recorded and stored. The app also logs rich process data which allow reconstructing the timeline of student interaction with the app, such as timestamps for the beginning and end of each student and narrator turn and answers to comprehension questions.

3 DATA

The data were collected in two summer camp programs during June-August of 2018.[3] One program ran for 6 weeks and included a reading session with the app for 20-50 minutes four days a week, with fewer days in the first week of the camp. The second program ran for the total of 8 weeks (different children were enrolled for a different number of weeks) with a variable reading schedule depending on other camp activities; each reading session included about half an hour of reading and half an hour of related games and activities. While other reading-based activities, unrelated to reading with the app, occurred during camp, neither of the camps specialized in reading instruction. In both camps, the children used the app to read J.K. Rowling's *Harry Potter and the Sorcerer's Stone* [12] and listen to the audiobook narration by Jim Dale [13]. The book contains about 80K words and the narration lasts 8.5 hours. Children read on mobile devices in a fairly laid back, informal atmosphere; see Figure 2.

The total of 36 children had at least one reading turn logged by the app, for a total of 2,390 logged turns. For the purposes of our analyses, we exclude children with <20 reading turns. This excludes 4 children and 51 turns. Of the remaining 32 children, 14 were boys and 18 girls, aged 8-11; mean age was 9 years 8 months as of end of camp. Table 1 describes the data used in this paper.

4 FROM DATA TO ANALYTICS

For this analysis, our target performance variable is percent correct in reading comprehension questions that measures how well children pay attention to important details of the plot of the story.

[2]https://readium.org

[3]We use the data slice from June 25 through August 13. One of the camps was still in session as of August 13. The cut-off was done to allow for processing and analysis.

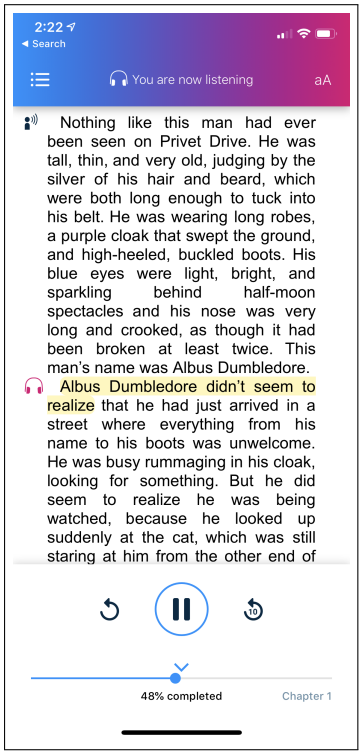

Figure 1: A screenshot of the iOS version of MyTurnToRead.

Figure 2: Children reading at one of the summer camps.

Table 1: Data description. The table shows the distributions of aggregated values for each child. N=32.

type of data	mean	std	min	max
# reading turns	73	42	23	174
# questions answered	63	38	20	153
% answered correctly	66	17	33	100
total audio time recorded (min)	87	68	7	302

Our goal is to explore relationships between this variable and other aspects of reading behavior measured by the app, based on reading research. In particular, reading research shows a strong relationship between reading comprehension and reading rate [15], which, in turn, is a part of the construct of oral reading fluency [11, 15, 18]. We seek to observe whether the relationship holds in our case.

4.1 Reading comprehension vs reading rate

Reading rate is usually measured as words per minute, or WPM. To calculate a student's reading rate for the given turn, we divide number of words in the text passage by the the recorded duration (in minutes). We observe a medium-to-strong *negative* correlation of r=-0.63 between average reading rate and percent correct in the reading comprehension questions, which seems to suggest, on the face of it, that children who read faster comprehend less well.

To further explore this first result, we examined more closely the relationship between reading rate and percent correct in comprehension questions; see Figure 3. The most striking feature of the plot is the scale of the estimated reading rate, reaching thousands of words per minute *on average across turns* for some readers. Such reading rates are highly implausible; the narrator of the audiobook, an accomplished reader and performer, reads at a rate of about 164 words per minute, on average, for this book. Given that our estimate is length of passage *to be read* in words over actual time spent in minutes, it stands to reason to assume that quite often the child *did not actually read the whole passage* before clicking "Done".

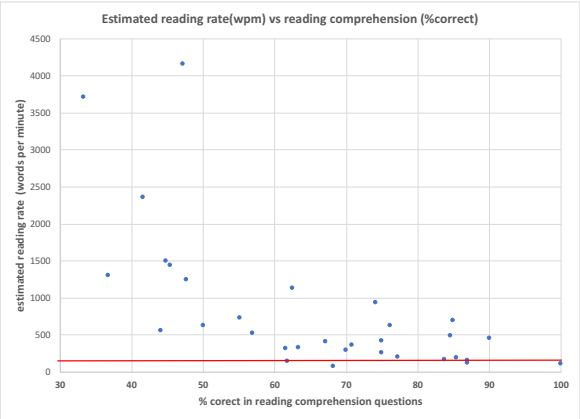

Figure 3: Average reading rate (calculated as length of passage over turn duration) vs. percent correct in reading comprehension questions for each student in the sample. The horizontal red line indicates average reading rate of the narrator of the book; clearly, values much above the red line are not plausible.

4.2 Estimating reasonable durations

Our next step is therefore to estimate boundaries for turn durations that at least *could* be a complete, bona-fide reading. This is possible since there are published oral reading fluency norms for elementary school children [2]. Thus, a 5th grader in the 90th percentile in the winter term would read 183 words per minute; a 3rd grader in the 10th percentile in the winter term is expected to read 62.[4]

[4]The norms are for words *correct* per minute (WCPM). The upper boundary for reasonable WPM would be higher than for WCPM, but not much higher, since the better (and faster) readers tend to read more accurately. We use the WCPM-based estimates.

The second piece of information needed for estimating a reasonable duration per passage is an estimate of within-person variation in reading rates across a large variety of passages. We are not aware of studies that would provide such estimates for children, but a recent study that tracked two proficient adult readers across hundreds of passages with varying textual characteristics has reported a normal-like distribution of reading rate across passages with a standard deviation of just under 10% of the average reading rate for that person [5].

We can thus set relatively permissive boundaries for reading rate per passage as being between a case of a slow reader who reads a passage that yields a reading rate that is two standard deviations below this reader's average (62 - 2*6 = 50) and a case of a fast reader who reads a passage that yields a reading rate that is two standard deviations above the average reading rate for that reader (183 + 2*18 = 219). Thus, we do not expect bona-fide reading to be done at a rate faster than 219 words per minute or slower than 50 words per minute. Using the actual number of words for every given passage, we thus have a duration range that *could* correspond to a bona-fide reading of the passage.

4.3 Re-analysis using reasonable durations

We then re-analyzed the data, taking only turns with reasonable durations (N=1,335 turns, or 57% of all turns), as the rest of the turns are unlikely to be complete readings. The strong negative correlation disappears; in fact, we find no correlation at all (r=-0.04). So, while the counter-intuitive result is explained by presence of turns that were not actually fully read, we still do not see the expected positive correlation between reading rate and comprehension.

One possible reason is that the subset of turns with reasonable durations might not represent faithfully a child's reading behavior. Thus, for one reader, out of the total of 128 reading turns, 126 are removed because their durations are shorter than the lower boundary of a reasonable duration. This reader's rate is thus estimated using only 2 turns, which could easily yield an unreliable estimate.

Secondly, our method of estimating reasonable durations is relatively permissive, so the subset of turns with reasonable durations might still include turns in which the child did not complete the expected reading but for one reason or another achieved the expected turn duration. Substantial presence of such turns in a child's data could undermine estimation of reading rate.

If we restrict the sample further to only children for whom reasonable turns constitute at least 70% of all their turns, that is, children for whom turns with reasonable duration dominate their reading behavior (N = 11 children), we observe the expected positive correlation between reading rate and comprehension (r = 0.53).

4.4 Proportion of reasonable turns (PRT) vs reading comprehension

As the reader might have noted, the subsetting to reasonable turns and further subsetting to children who tend to have reasonable turns resulted in a dramatic reduction in the size of the dataset. Indeed, as many as 43% of the turns are outside of the reasonable bounds, where 36% are too fast and 7% are too slow. The distribution of proportion of reasonable turns (**PRT**) per reader has a mean of

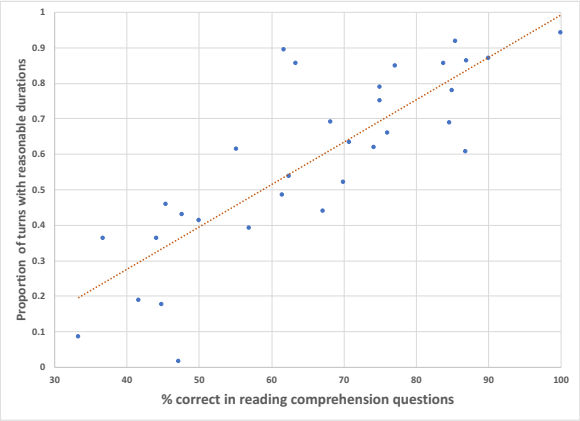

Figure 4: PRT (y-axis) vs % correct in comprehension questions (x-axis); N=32.

0.59 and standard deviation of 0.25. This suggests that proportion of likely bona-fide turns shows large variation across readers.

Our last, exploratory step, was to check whether the observed variation in PRT for each child corresponded to the target performance variable, percent correct in comprehension questions. We found the two variables to be very highly correlated: r = 0.83. Figure 4 shows the plot relating the two variables.

5 DISCUSSION

We identified a novel measure of independent reading behavior in the context of extended interactive reading that is highly predictive of comprehension performance. This close relationship is easy to explain – if you tend to *not* read your turns completely, you do not get all the information needed to respond to comprehension questions. We think about it as a tendency to take advantage of reading opportunities, or to "show up": In order to make success at reading comprehension even possible, you need to actually read. Success in reading comprehension (as far as it is evidenced by percent correct in surface comprehension questions) is apparently directly related to *consistently acting on opportunities to read.*

One potential caveat to this interpretation of the measure is that we do not know that within-bounds turns actually correspond to a reasonable attempt at reading. In principle, it is possible that the child was staring at the ceiling but knew when to stop staring just so the duration of his staring would coincide with reasonable boundaries of reading duration. To check whether students with higher showing-up tendency did in fact tend to read more, we ran all the recordings of the children's reading turns through automated speech recognition (ASR) [6]. We then used the ASR hypothesis to compute the proportion of words in the passage that the system recognized as correctly read by the child. Finally, we computed average per child and correlated that with the child's tendency to "show up". We observed a correlation of r = 0.77, suggesting that children who tended to "show up" also tended to read, not just stare at the screen for the right amount of time.[5]

Another potential objection to thinking about PRT as a tendency to actually read the text is that perhaps children who had a low PRT were still reading – silently, instead of orally, as instructed. While this possibility cannot be completely ruled out, the following two observations suggest that this is not likely to be a substantial occurrence in our data. First, [17] put the average rate of silent reading in 5th grade at 173 wpm and at 158 for 4th graders; [3] replicated the latter, with 154 wpm. Thus, to read consistently at more than 219 wpm would take a very fast reader, for this age; it is unlikely that many children in our sample could sustain such a rate. Secondly, Figure 4 suggests that readers with a high proportion of unreasonably fast (putative) readings did poorly on comprehension questions. It is unclear whether moving through the text very fast without taking in sufficient information to respond to questions should count as "silent reading". Indeed, in [3], very fast readers who could not respond to comprehension questions were removed from the norming study of silent reading rate. Our data do not seem to have a substantial presence of extremely fast, silent, comprehension-effective readings.

Children with low "showing up" seem to have consistently refused to read. It is possible that they perceived the reading as either requiring too much effort or even perhaps insurmountable. In consultation with the instructors in one of the camps, we reduced the average duration of student turns from 150 to 50 words (while keeping the narrator turn at 200 words) for two children with extremely low "showing up", as of half way into the program. The intervention was ineffective for one of them: [%"showing up", %correct] were [16%,45%] before the change and [20%,45%] after, but looked more promising for the other child: [23%,44%] before vs [57%,67%] after. Continuously monitoring the "showing-up" analytic may create opportunities for individualized assistance such as reduction in task difficulty; more experiments are needed for best time-point and extent of such interventions.

6 RELATED WORK

6.1 Measurement of effort in assessment

Low effort might interfere with demonstration, and therefore with assessment, of true ability. In addition, unmotivated responses could hamper estimation of item difficulty. These considerations led to a recommendation of "motivation filtering" in the educational measurement literature [19], namely, removal of low-motivation responses from the data. Solutions for detecting low-motivation responses include self-report of effort [16], as well as automatic measurements based on plausibility of response time [20]. Response Time Effort (RTE) is the proportion of items for which the response was obtained too quickly for the test taker to actually have read the question [20]. In [20], RTE is not correlated with measures of academic ability; however, what is true of a one-time low-stakes testing

[5]Words from the passage detected in the speech signal are strong evidence that the child has read, but non-detection is *not* strong evidence of the opposite; a child could have read silently, or even orally, but due to noise, interference from another reader nearby, or to muttering, ASR did not detect what was said [6]. Duration measurement is more robust than ASR but is more permissive, giving readers the benefit of the doubt regarding whether or not they actually read, as long as the duration was reasonable.

situation with adults might not hold in a context of children's reading across an extended period of time: the not-showing-up tendency *repeated every day* could be a whim but could also reflect reluctance due to anticipated difficulty. Better understanding of reasons for low showing-up behavior is a major goal for our future work.

6.2 Reading development in children

Reading continuous text accurately, at a grade-appropriate rate, with good prosody and expression is evidence that foundational reading skills (decoding, word recognition, vocabulary, and phrasal and syntactic processing) are developing along expected trajectories [15]. Currently, oral reading probes are a commonly used instrument in the early grades for monitoring progress as indicators of reading fluency and of comprehension [18]. However, the time and effort to collect this information are considerable, so deriving it unobtrusively and continuously through an app used for extended reading would be helpful for learners and teachers. The techniques applied in this study represent steps along that path.

6.3 Automated support of oral reading

Automated reading tutors have been around for a while. Project LISTEN [8] is one of the most mature systems in this area. [1] provide an overview of some of more recent developments in the area of technology-based literacy instruction. A unique challenge in our application is the context of extended reading sessions in a relaxed laid-back atmosphere over a period of several weeks which makes it difficult to consistently control the environment. However, with the growth of speech-centric devices, noise robustness is becoming a core ASR technology, dealing with speech collected in relatively adverse conditions [4]. Connecting automated processing of the child's oral reading to the app is future work.

7 CONCLUSION

It is difficult to overstate the importance of literacy for adequate functioning in society, from educational attainment and employment opportunities to health outcomes [14]. We created a reading app with the goal of helping readers improve their reading skill, and used it to collect unique data on reading behavior when reading a long book for an extended period of time in a relaxed, low-stakes context. The collected data reflect behaviors that would traditionally be filtered out as "noise". Yet we show that aggregating such information across an extended period of time yields a consistent *behavioral* pattern which is strongly predictive of performance.

In the context of using analytics to obtain new knowledge about the process of reading in general and about the interplay between speed of reading and comprehension specifically, our findings suggest that actually getting down to the reading *consistently across many such opportunities* is an important mediating factor. Thus, once the reader is in fact acting on an opportunity to read in a consistent fashion, the expected relationships between reading subskills are observed. Inconsistent reading behavior in which many reading opportunities are not fully utilized leads to a breakdown in the observed subskill relationship and in the reading comprehension itself. The visionary author of children's books was thus quite prescient by placing *could you* right next to *would you*, and not only for eating green eggs.

8 ACKNOWLEDGEMENTS

We would like to thank K. Dreier, V. Liceralde, and J. Bruno for creating the questions; B. Gyawali and P. Lange for text-and-audio alignment of the book; Y. Qian and A. Misra for help with ASR; Y. Donev, G. Angelov and the rest of Astea Solutions team for the outstanding app design and development work; site administrators and instructors in the two summer camps for their interest and enthusiasm implementing the summer reading program with MyTurnToRead.

REFERENCES

[1] S. Crossley and D. McNamara. 2016. Educational Technologies and Literacy Development. In *Adaptive Educational technologies for literacy instruction*, Scott A. Crossley and Danielle Mcnamara (Eds.). Routledge, New York, 1–12.
[2] J. Hasbrouck and G. Tindal. 2017. An update to compiled ORF norms. Technical Report No. 1702. http://www.brtprojects.org/wp-content/uploads/2017/10/TechRpt_1702ORFNorms_Fini.pdf
[3] E. Hiebert, K. Wilson, and G. Trainin. 2011. Are students really reading in independent reading contexts? An examination of comprehension-based silent reading rate. In *Revisiting silent reading: New directions for teachers and researchers*, E. Hiebert and D. Reutzel (Eds.). International Reading Association, Newark, DE.
[4] J. Li, L. Deng, Y. Gong, and R. Haeb-Umbach. 2014. An Overview of Noise-Robust Automatic Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22, 4 (April 2014), 745–777.
[5] A. Loukina, V. Liceralde, and B. Beigman Klebanov. 2018. Towards Understanding Text Factors in Oral Reading. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, New Orleans, Louisiana, 2143–2154.
[6] A. Loukina, N. Madnani, B. Beigman Klebanov, A. Misra, G. Angelov, and O. Todic. 2018. Evaluating on-device ASR of field recordings from an interactive reading companion. In *Proceedings of SLT2018*. IEEE, Athens, Greece.
[7] E. Meisinger, B. Bradley, P. Schwanenflugel, and M. Kuhn. 2010. Teachers' Perceptions of Word Callers and Related Literacy Concepts. *School Psychology Review* 39, 1 (2010), 54–68.
[8] J. Mostow. 2012. Why and how our automated reading tutor listens. *Proceedings of the International Symposium on Automatic Detection of Errors in Pronunciation Training* (2012), 43–52.
[9] National Assessment of Educational Progress. 2017. NAEP Reading Report Card: National achievement level results. https://www.nationsreportcard.gov/reading_2017/#/nation/achievement?grade=4.
[10] University of Phoenix College of Education. 2017. K-12 Edtech and homework in the classroom survey. https://www.phoenix.edu/about_us/media-center/news/majority-of-teachers-think-homework-is-beneficial.html.
[11] Y. Petscher and Y.-S. Kim. 2011. The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology* 49, 1 (2011), 107–129.
[12] J.K. Rowling. 2015. *Harry Potter and the Sorcerer's Stone.* Published as EPUB file by Pottermore Ltd.
[13] J.K. Rowling and Jim Dale. 2016. *Harry Potter and the Sorcerer's Stone.* Listening Library/Penguin Random House.
[14] J. Sabatini, T. O'Reilly, and N. Doorey. 2018. Retooling literacy education for the twenty-first century: Key findings of the reading for understanding initiative and their implications. Princeton, NJ: Center for Research on Human Capital and Education, Educational Testing Service and CCSSO.
[15] J. Sabatini, T. O'Reilly, and Z. Wang. 2018. Relating reading comprehension to oral reading performance in the NAEP fourth-grade special study of oral reading. *Reading Research Quarterly* (2018). https://doi.org/10.1002/rrq.226
[16] P. Swerdzewski, C. Harmes, and S. Finney. 2011. Two Approaches for Identifying Low-Motivated Students in a Low-Stakes Assessment Context. *Applied Measurement in Education* 24, 2 (2011), 162–188.
[17] S. Taylor, H. Frankenpohl, and J. Pettee. 1960. Grade level norms for the components of the fundamental reading skills. *EDL Research and Information Bulletin* 3 (1960).
[18] M. Wayman, T. Wallace, H. Wiley, R. Ticha, and C. Espin. 2007. Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education* 41, 2 (2007), 85–120.
[19] S. Wise and C. DeMars. 2010. Examinee Noneffort and the Validity of Program Assessment Results. *Educational Assessment* 15, 1 (2010), 27–41.
[20] S. Wise and X. Kong. 2005. Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education* 18, 2 (2005), 163–183.