Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID

Publication/Completion Date—(if *In Press,* enter year accepted or completed)

Check type of content being submitted and complete one of the following in the box below:
- ○ If article: Name of journal, volume, and issue number if available
- ○ If paper: Name of conference, date of conference, and place of conference
- ○ If book chapter: Title of book, page range, publisher name and location
- ○ If book: Publisher name and location
- ○ If dissertation: Name of institution, type of degree, and department granting degree

DOI or URL to published work (if available)

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** through **[Grant number]** to **Institution]** .The opinions expressed are those of the authors and do not represent views of the **[Office name]** or the U.S. Department of Education.

How Item and Learner Characteristics Matter in Intelligent Tutoring Systems Data

John Hollander[(✉)] , John Sabatini , and Art Graesser

University of Memphis, Memphis, TN 38111, USA
`jmhllndr@memphis.edu`

Abstract. AutoTutor-ARC (adult reading comprehension) is an intelligent tutoring system that uses conversational agents to help adult learners improve their comprehension skills. However, in such a system, not all lessons and items optimally serve the same purposes. In this paper, we describe a method for classifying items that are *instructive, evaluative, motivational*, versus *potentially flawed* based on analyses of items' psychometric properties. Further, there is no a priori way of determining which lessons are optimal given the learner's reading profile needs. To address this, we evaluate how assessing learner component reading skills can inform various aspects of learner needs on AutoTutor lessons. More specifically, we compare learners who were classified as *proficient*, *underengaged*, *conscientious*, versus *struggling* readers based on their experiences with AutoTutor. Together, these analyses suggest the utility of integrating assessments with instruction: efficient, adaptive learning at the lesson level, more efficient and valid posttesting, and consequently, recommendations for more targeted, adaptive pathways through the instructional program/system.

Keywords: Intelligent tutoring systems · Reading skills · Psychometrics

1 Introduction

1.1 Adaptive Education and Adult Literacy

Assessments of worldwide literacy rates indicate that around 14% of adults may be classified as low literate [1]. While advances in research and technology are helping more adults improve their ability to read and write than ever, the best efforts of educators and literacy researchers still do not meaningfully help a significant portion of this population. The development of adaptive learning technologies could significantly address this problem because adult learners are a dispersed and diverse population [2–4]. We analyzed data obtained from adults with low literacy who completed a reading component skill assessment battery before and after participating in an instructional program using AutoTutor–ARC, an adult literacy-focused intelligent tutoring system with two conversational agents that periodically ask the learners questions while adults read texts and other learning materials. The lessons are specifically designed to engage adult learners, and range from word-level learning to practical applications of complex literacy skills.

However, determining how AutoTutor-ARC lessons and items relate to specific reading component skills known to impede or facilitate comprehension growth would allow for a more responsive and effective approach [5, 6].

The Reading Inventory and Scholastic Evaluation (RISE) (also known as the Study Aid and Reading Assessment or SARA) is a battery of six reading component skills subtests measuring skills that are known to be malleable to instruction [7], specifically *decoding and word recognition*, *vocabulary*, *morphology*, *sentence processing*, *reading efficiency*, and *reading comprehension.* A reader who lacks adequate component skills may rely on one or more compensatory behaviors, strategies that are often not optimal to continued growth [3].

2   Method

Data used for this study were obtained from three waves of an adult literacy intervention study. Participants included 252 adult literacy learners ($M_{age} = 42.4$, $SD = 13.9$, 74.6% female), who were offered 100 h of instruction (featuring hybrid classes of teacher-led sessions and AutoTutor sessions) over the course of four months. Auto-Tutor lessons were assigned to students individually by their teacher; not all students took all lessons, and lessons could be repeated. Participants completed one form of the RISE before the intervention, and another form afterward [7].

We gauged how well ITS data can be used to identify learner characteristics with respect to adult literacy by adopting the results of Fang et al.'s 4-cluster clustering analysis [5] of adult learners using AutoTutor-ARC. These clusters were defined by their accuracy and speed in answering conversation-based questions during learning: *proficient readers* (accurate and fast), *struggling readers* (inaccurate and slow), *conscientious readers* (accurate but slow), and *underengaged readers* (less accurate but fast).

3   Results

We considered items in lessons which were fully completed by at least 90 participants on their first attempt. In accordance with similar analyses and data processing procedures [5], we considered items that were correctly answered by at least 95% of participants to be *motivational*, as they do not provide any new information about learner knowledge. Further, we considered items with a negative item-total correlation to be *potentially flawed*. These items are psychometrically inconsistent with lesson topic constructs. Figure 1 contains a graphical representation of this classification.

We calculated the reliability of each of these lessons, once with all items, and once with *potentially flawed* items removed. Three items were removed from the Text Signals lesson, increasing its reliability from α = .470 to α = .550, while its average item accuracy (74%) remained the same. Five items were removed from Word Parts, increasing its reliability from α = .307 to α = .62, decreasing its average accuracy (66% to 62%). One item was removed from Main Ideas, increasing its reliability from α = .279 to α = .340 with no effect on its average accuracy (67%).
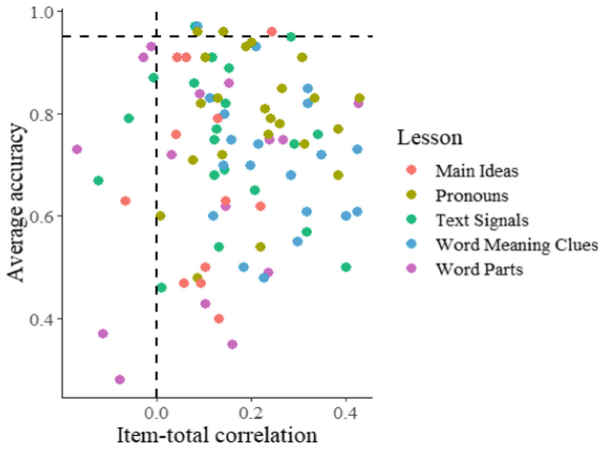
Fig. 1. Item characteristics by lesson. Items to the left of the vertical dashed line are considered *potentially flawed.* Items above the horizontal dashed line are considered *motivational.*

We then created a list of the RISE subtests whose constructs are theoretically aligned with the AutoTutor lessons. We concatenated the item-level data of each lesson-subtest pair. In cases in which more than one subtest was theoretically aligned, we created two separate sets. We then calculated the item-total correlation of each AutoTutor lesson item within its respective lesson-subtest combination. We labeled AutoTutor items with item-total correlations greater than .15 in each pairwise combination as *evaluative items.* We then calculated the reliability of these sets containing RISE subtests plus their *evaluative* AutoTutor items (see Table 1). The remaining unlabeled items were considered *instructive items* because they may have value for learning, but do not map onto the assessment-based constructs in question.

Table 1. Cronbach's ɑ for RISE subtests when combined with each AutoTutor lesson pairwise (baseline RISE reliabilities displayed parenthetically).

	MORPH (ɑ = .889)	SEN (ɑ = .818)	RC (ɑ = .539)
Text signals	0.896 (40%)	0.883 (40%)	
Word parts	0.899 (53%)		
Word meaning clues			0.629 (55%)
Pronouns			0.627 (39%)
Main ideas		0.826 (22%)	0.57 (6%)

4   Discussion

Within the domain of adult literacy education, we have provided examples of how independently developed assessments and ITSs can inform one another to better account for how the characteristics of students and lesson items intersect.

We examined how item characteristics may be leveraged to further integrate assessment and instruction, using industry-standard psychometric analytic techniques, to align item properties to independently valid subtests. We created a taxonomy of lesson items as: *potentially flawed* (psychometrically inconsistent with lesson topic constructs), *motivational* (generally too easy to be informative), *evaluative* (closely related to assessment-oriented skill/knowledge constructs), and *instructive* (consistent with lesson topics, but not external constructs). In support of the validity of the resulting taxonomy, we found that including evaluative items in reliability analyses of construct aligned subtests improved the reliability, supporting the generalization of item-level performance during instruction to specific, psychometrically validated frameworks; analogously, removing flawed items increased reliability of the remaining lesson items. Thus, this taxonomy and analytic frame can be useful to adaptive systems by enhancing assessment precision and instructional content validity.

Future research should more closely examine the most effective use of item and lesson characteristics in real-time ITS, to adapt learning activities and estimate student proficiency as learning progresses. Future research should explore how items embedded in assessments versus learning environments may interact with learner profiles, perhaps predicting which content will be frustrating or challenging to different learners.

References

1. UNESCO: Literacy Rates Continue to Rise from One Generation to the Next. UIS Fact Sheet No, 45. (2017)
2. Greenberg, D.: The Challenges Facing Adult Literacy Programs. Community Lit. J. **3** (2008). https://doi.org/10.25148/clj.3.1.009480
3. Sabatini, J., O'Reilly, T., Dreier, K., Wang, Z.: Cognitive processing challenges associated with low literacy in adults. In: The Wiley Handbook of Adult Literacy, pp. 15–39. Wiley (2019). https://doi.org/10.1002/9781119261407.ch1
4. Comings, J.P., Soricone, L.: Adult literacy research: opportunities and challenges. Boston, MA (2007)
5. Fang, Y., et al.: Patterns of adults with low literacy skills interacting with an intelligent tutoring system. Int. J. Artif. Intell. Educ. 1–26 (2021). https://doi.org/10.1007/s40593-021-00266-y
6. Chen, S., et al.: Automated disengagement tracking within an intelligent tutoring system. Front. Artif. Intell. **3** (2021). https://doi.org/10.3389/frai.2020.595627
7. Sabatini, J., Weeks, J., O'Reilly, T., Bruce, K., Steinberg, J., Chao, S.F.: SARA Reading Components Tests, RISE Forms: Technical Adequacy and Test Design, 3rd edn. ETS Research Report Series 2019, pp. 1–30 (2019). https://doi.org/10.1002/ets2.12269