WriteToLearn®

General Overview of WriteToLearn™ and Its Components

Goals, foundations, instructional and formative assessment functions, summary of evidence for reliability, validity and interventional effectiveness

PEARSON

Motivating Educational Philosophies

WriteToLearn integrates practice and assessment in reading comprehension with writing about what is learned, a natural symbiotic combination for literacy learning. Its personalization, game-like improvement-scoring and individualized progress monitoring were designed and iteratively refined to make its use enjoyable and motivating for students. Its real-time and longterm reports about student activity and progress were designed and refined to give teachers and schools information that is useful and usable for augmenting and guiding classroom instruction and curricular decisions. Another guiding goal was to embed assessments in actual performance of reading to learn and writing to express learned knowledge. At Pearson, we believe that this makes assessment better able to promote abilities that will be used in and out of the classroom than are traditional tests of indexical tasks such as question answering. Therefore, we believe, that such embedded assessment should also produce more valid assessment.

The design of WriteToLearn was also influenced by three widely endorsed and well research-supported principles. One is that most learning to read and write well comes from reading and writing, the more the better. The second is the widespread use, opinion and evidence in favor of summarization as the best single learning activity to foster comprehension skills. Third, is the great value of making evaluative and instructive feedback come immediately after and tied to the actions it is intended to reinforce or correct. Thus, it seems to us that the best formative assessments must be ones that encourage, instruct and reward progress in reading and writing while it is happening.

Research Background of WriteToLearn

WriteToLearn's most important technological basis is Latent Semantic Analysis (LSA), a statistical language learning theory and computer model that measures the semantic similarity of words and documents with accuracy closely approximating that of human judges. LSA originated at Bell Laboratories under Tom Landauer; its translation into an educational technology component took place at the University of Colorado (CU) and Pearson's Knowledge Technologies group.

In addition to LSA, WriteToLearn employs other powerful new statistical classification and analysis methods, as well as techniques from computational linguistics, spell- and grammar-correctors, and item-response theory.

WriteToLearn is based on ten years of research and evaluation at the University of Colorado and New Mexico State University, five as part of an IERI research and effectiveness trial project at CU, combined with seven years of professional educational software development and both software and educational effectiveness testing at Knowledge Analysis Technologies (since 2004, Pearson's Knowledge Technologies group.) At CU, the research was under the direction of Profs. Walter Kintsch and Tom Landauer, at NMSU under Prof. Peter Foltz, the latter two are now also in charge of research at Pearson's Knowledge Technologies group. Researchers Prof. Louis Gomez at Northwestern University School of Education and Social Policy and Dr. Jack Stenner of MetaMetrics, Inc have also collaborated in the research behind WriteToLearn.

About the Summary Street® Component of WriteToLearn

Introduction

The Nation's Report Card for Reading[1] estimates that about half of eighthand twelfth-grade students score at or below a basic level of reading comprehension. In addition, according to the National Reading Panel, children can read [in the sense of decoding words from print to sound], but they often don't understand the meaning of what they're reading or how to appreciate what's relevant[2], which can hobble comprehension and enjoyment and lead not only to poor test scores but inability to read schoolwork with understanding. Summarizing what has been read is not only an effective strategy for increasing its comprehension at the time, but helps build lasting ability to comprehend.

Typically, summarizing is introduced to students in the third grade, but it continues as a part of instruction into higher grades as well. Unfortunately however, the time needed for teachers (who may teach as many as 300

different students each week) to evaluate and critique individual summaries severely limits their potential benefits. To create more opportunities for students to engage in evaluated summarizing than teachers have time for, University of Colorado at Boulder researchers and Pearson's Knowledge Technologies group developed Summary Street. Summary Street is one of the two paired components of WriteToLearn (the other being the Intelligent Essay Assessor, which expands opportunities to write and receive tutorial feedback on open-ended essays in response to topical prompts.) Summary Street is an automated, web-based tool that evaluates and critiques both the substantive content of students' summaries and the way they are written, and provides helpful feedback on how to improve on successive revisions. Summary Street requires students to express the main ideas of what they have read in many fewer and their own words, typically 1/5 as many. It measures the student's section-by section content coverage by comparing it with the original reading using a powerful computer model (LSA) that can assess whether or not the right information is conveyed even if different words, phrases or organization is employed. The division into sections is accomplished either semi-automatically on the basis of the text format or by an author or editor. WriteToLearn also gives advice on spelling, correction of clear and reliably assessed grammar errors, and redundant and irrelevant sentences.

Obviously, Summary Street does not do everything skilled teachers could do, nor do what they do as well, but it does provides some of the same tutorial benefits. Most importantly, it amplifies opportunities and motivation to read with understanding and to express accurately what one learned by reading. Observation and surveys have shown that students enjoy working with Summary Street where they often find ordinary summary writing assignments burdensome and boring.

With Summary Street, teachers can select texts from a potentially unlimited variety of content areas, currently, for example, ranging over more than 300 science, social studies, history and age-appropriate biography and fiction texts, including selected readings from Pearson's Prentice Hall middle grades textbook programs, "Science Explorer," "World Studies," and Scott Foresman Reading Street leveled readers. The basic catalog of reading texts

spans the grades 4-12 curriculum and a complete range of difficulty as measured by Lexiles and other approaches. Additional texts are added on a regular basis. The passing thresholds toward which students work in improving their summaries are initially set by a computer algorithm, but are typically adjusted by teachers to suit their pedagogical goals and the individual needs of their students.

On their own laptops, teachers can view, in real time, automated, largely graphical reports and visualizations that display individual student and class activity and progress. WriteToLearn also gives teachers immediate access to all successive revisions of any student's summaries and the automatic content comprehension and writing quality feedback they received. In our view, this kind of formative assessment gives ELA (and subject matter) teachers not only a powerful new tool for intervening at the right times with the right help, but also places a set of progress measurements where they can be unusually focused and learning enhancing.

Cumulated results of student and class activities and progress—what was read, score means, distributions, criteria attainment—on content comprehension, knowledge expression and writing traits—kept by Summary Street (and IEA as well) also constitute an unusually rich source of face-valid formative and summative assessments for literacy skills. With the right choice of readings by teachers or alignment with readings from systematic reading programs or state standards (Pearson Ed can also provide the former, PKT the latter), Summary Street can, and has been, used effectively across the entire curriculum and from grades 4 to 12 (and to a limited extent in higher education, adult, ELL and professional training.)

The reliability of Summary Street's individual feedback scores has received less direct evaluation so far than other components of WriteToLearn. However, several evaluation studies using double blind teacher ratings have shown adequate levels of reliability and evidence of rapid and significant learning effects relative to randomly assigned control groups (reported in detail later). Taken together with validity evidence from school-grade associated performance, strong effects on some state test reading comprehension item scores, and system and blind rating-measured progress compared to controls in true experiments. There is, thus, little doubt that

the reliability is sufficient for the motivation and learning promotion goals of the intervention.

To date, more than 20,000 students have used Summary Street. In addition to significant improvement in student writing and comprehension performance across subject areas, students report that they find using Summary Street motivating, rewarding and fun, and teachers have greeted it with virtually unanimous enthusiasm.

How Summary Street Works

The Summary Street component of WriteToLearn consists of a student interface, a teacher interface and the Knowledge Analysis Technologies™ (KAT) engine, which automatically evaluates the semantic substance of texts by analyzing a passage as a holistic meaning, not by looking for particular words. In its essay-grading applications, as detailed later, the technology has been found to evaluate writing as accurately as skilled human graders, usually correlating with each two graders better than they correlate with each other. Whether it does that as well in its Summary Street application has yet to be definitively determined, but its adequacy is attested by high correlations with blind teacher ratings on overall quality and writing traits, and by its demonstrated positive effects on learning.

Student Interaction: Students initiate interaction with the system by logging in using a web browser. They select a text to summarize from a library of materials, enter their summaries through a simple editing window, and receive immediate feedback on qualities of their summaries. The figure below shows the graphic feedback that allows students to pinpoint and address writing difficulties. With each subsequent change to their summaries, students can track their progress toward the goals specified by their teacher via Summary Street.
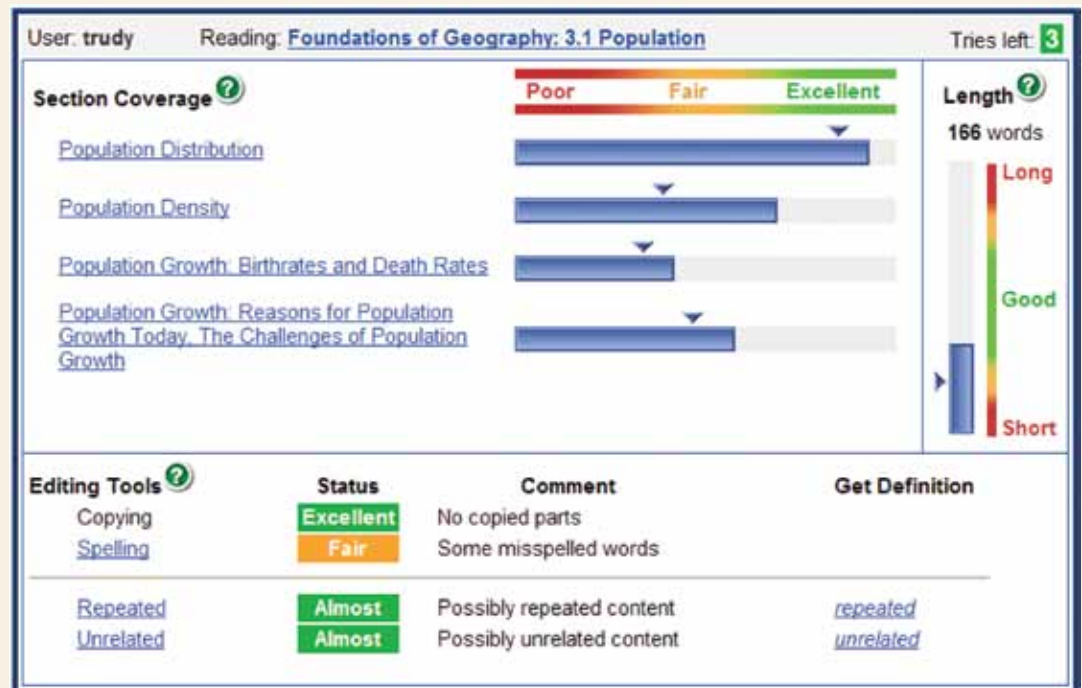
WriteToLearn®

User: trudy   Reading: Foundations of Geography: 3.1 Population   Tries left: 3

Section Coverage ❓   Poor   Fair   Excellent   **Length** ❓
                                                                        166 words

Population Distribution

Population Density

Population Growth: Birthrates and Death Rates

Population Growth: Reasons for Population Growth Today, The Challenges of Population Growth

Long

Good

Short

Editing Tools ❓	Status	Comment	Get Definition
Copying	Excellent	No copied parts	
Spelling	Fair	Some misspelled words	
Repeated	Almost	Possibly repeated content	repeated
Unrelated	Almost	Possibly unrelated content	unrelated

Figure 1: *Summary Street Feedback Screen. This scoreboard is presented immediately after the student clicks "Get Feedback."*

Summary Street provides feedback on whether:

- The summary covers the key points of each section of the reading passage
- The summary has been condensed from the original text (the general guideline is about 15-20 percent of the original text length)
- Students have used their own words sufficiently, or copied too much from the original
- There is too much repetition (detected even if words or phrases are used)
- There are unrelated sentences that don't add anything to the overall value or meaning of the summary and can be omitted or combined
- There are spelling or other errors for which it offers assistance.

Teacher Interaction. Teachers log into WriteToLearn to set up class rosters, administer assignments and monitor activity and progress. Student proficiency levels can be addressed by assigning materials at different difficulty levels, by modifying content scoring thresholds and expected summary length, and by providing different spell-checking options. Teachers can also monitor class and student performance by choosing from a number of reports, starting at the class overview level or drilling down into the successive summary revisions of particular students.

Class: Demo Writing 2
School: Demonstration School
Report Date: Thu Oct 4 2007, 4.02 pm
Help

Ancient Civilizations -- Aztec

Sections:
1. Territory
2. Agriculture and Trade
3. Aztec Lifestyle and Beliefs
4. Artistic and Scientific Contributions
5. The Spanish Conquest

Preferred Length: 200 - 350 words

Student Name (Login)	Section 1	Section 2	Section 3	Section 4	Section 5	Word Count	Copying %	Spelling Errors	Repeated %	Unrelated %	Minutes on Task*	Attempts	Passing Attempts
Bizzy, Luke (luke)	Excellent	Good	Excellent	Fair	Fair	339	0	0	0	4	30	2	0
Brown, Sally (sally)	Excellent	Poor	Excellent	Poor	Fair	117	9	0	0	11	31	3	0
Furlong, Terry (terry)	Good	Excellent	Good	Fair	Fair	268	0	0	0	0	30	3	0
Lightoff, Tanya (tanya)	Fair	Excellent	Excellent	Fair	Fair	131	29	0	0	20	17	2	0
Marx, Audrey (audrey)	Excellent	Excellent	Excellent	Good	Good	234	0	0	0	0	34	4	0
Smith, Sam (sam1)	No Summaries												
Spelling, Erin (erin)	Fair	Fair	Excellent	Fair	Fair	266	0	15	0	0	34	3	0
Storie, Rita (rita)	Excellent	Excellent	Excellent	Excellent	Excellent	280	0	0	0	0	35	4	1
Woods, Trudy (_demotrudy)	No Summaries												
Averages Students: 9 With attempts: 7	Good	Good	Excellent	Fair	Fair	234	5.4	2.1	0.0	5.0	30.1	3.0	0.1

* -Time, in minutes, between feedback requests. Requests with intervals greater than one hour not included

Figure 2: *Teacher Class Overview Report. Indicates individual student and class performance on an activity.*

Why Summary Street Works

In the current educational environment, students, especially struggling ones, rarely have enough opportunities to practice writing skills with useful feedback either in the classroom or out. As research shows (e.g., Graham & Harris, 2005; Patthey-Chavez, Matsumura & Valdes, 2004), students benefit most from specific, immediate and individualized feedback on their performance, especially when it addresses content as well as surface-level features. Good feedback allows them to concentrate on particular deficits and improve their performance until they can meet pre-defined criteria.

WriteToLearn®

Providing an environment with frequent and rich feedback opportunities is beyond the limits of typical classroom instruction and often beyond the limits of supportive parents or tutors. The Summary Street component of WriteToLearn offers this kind of environment. It alerts students to specific problems in their summaries without penalizing or overcorrecting them. It provides a natural way to improve performance until learning criteria are met, and sets learning criteria that provide authentic indicators of proficiency. Research on the use of Summary Street has confirmed that this type of guided practice with immediate feedback helps students improve reading comprehension and writing skills.

About the Intelligent Essay Assessor™ Component of WriteToLearn

The Intelligent Essay Assessor (IEA) component of WriteToLearn automatically evaluates students' essays written to its wide range of carefully selected writing topics. Students enter their essays into an onscreen textbox and receive immediate holistic feedback as well as feedback on any set of more specific writing traits that is desired, and for which the system has been successfully trained and calibrated. Training and calibration customarily requires obtaining 300 or more representative essays that have been human scored, ideally by two independent expert readers. Several sets of traits that are as scorable by IEA as by humans have already been trained and are available for use in WriteToLearn. The figure below shows the graphic feedback that allows students to pinpoint and address their essay writing difficulties.

WriteToLearn scores on 6 popular traits of writing. Feedback is also provided on more mechanical aspects of writing such as grammar and spelling. Uniquely, through the use of the KAT engine, WriteToLearn can also evaluate redundancy, relevance and semantic coherence.
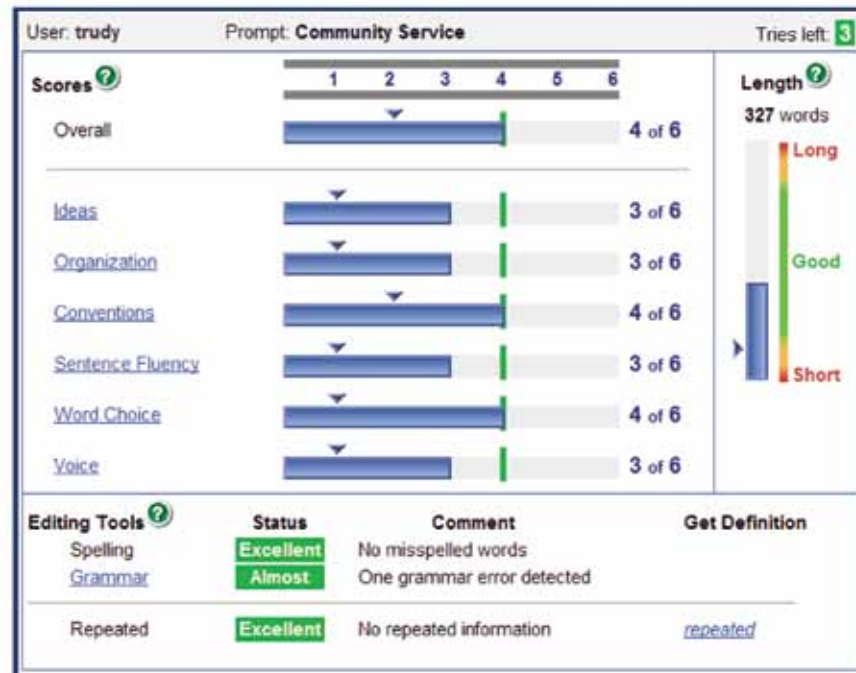
Figure 3: *Intelligent Essay Assessor Student Feedback Screen. Offers feedback that includes an overall holistic score, 6 traits of writing, as well as spelling, grammar and redundancy.*

IEA Reliability

Among the dozens of testing programs in which IEA has been used or evaluated and the several million essays that have been scored with it are some of special relevance to potential users of WriteToLearn, and to formative assessment goals. These are essays either written to prompts that are currently available for use, or of the type most suitable for WriteToLearn, or that have yielded extensive and accurate reliability and/or validity data or other information of special interest.

In one—to our knowledge unique—evaluation study conducted by an independent testing organization, 3,000 essays were written by 4th to 12th grade students to 12 different prompts in a specially constructed experimental design in which each student wrote on six essay prompts,

one on each of six testing days, each essay was scored independently by four different readers, and prompts, day, reader and day occurred in every possible combination.

This made it possible to evaluate fundamental scoring reliability much more accurately than ever before by statistically holding constant all the variables except whether the essays were scored by humans or IEA. The answer was that IEA correlated with the human readers significantly better than they correlated with each other.

Moreover, half of each set of prompts were used with students in two different school grades, each pair of grades two years apart. This allowed us to compare how well human and IEA scores measured average progress over two years of schooling. The answer: a dead heat.

In another large study, Pearson's Knowledge Technologies group used essays on 81 different reading-related topics answered by students in grades 6 through 12 online in a web-based companion to a Pearson's Prentice Hall reading series, a very similar application as in WriteToLearn. The correlation between IEA and Humans was better than that between the two human readers for every grade level, a very highly significant difference, by an average of .037, with probability less than one in a thousand. The proportion of exact agreements on the 6 point scale were nearly identical, 61.1% for IEA to human and 61.7% for human to human. Exact plus adjacent agreements were above 98% for both.

There have been many other studies of IEA reliability, some large some small. IEA has been successfully used to assess essays in a wide variety of academic, professional and employment training domains, ranging from story and letter writing through biology and history to military leadership to medical patient interviews to the ACT, SAT, TOEFL and GMAT and legal knowledge tests.

Other kinds of evaluation have included using early versions of the technology to test knowledge before and after reading a text. In one such study, a section from university biology textbook was read by high school students, college students and medical students. Both the technology and a traditional test were used to measure how much was learned as a result. The technology showed improvements for all readers, most for the high school students, least for medical students (which we called the Goldilocks effect) and correlated highly with the traditional test.

In a variety of experimental psychology laboratory studies, LSA, the heart of IEA, has been used successfully by many investigators to explore and predict the effects of similarity of meaning of words and phrases on short term and long-term memory. They have shown, for example, that difference in meaning similarity measurable by LSA but not apparent to human judges have significant effects on memory for words and text.

End Notes:

[1] National Center for Education Statistics, 2005.

[2] Donovan, M., & Pellegrino, J. (EDS), "Learning and Instruction, A SERP Research Agenda." Published by The National Research Council of the National Academies, Wash. D.C., p. 52.