



Diagnostic **A**sessment & **AC**ollege **S**

Developing and Validating the DAACS Writing Assessment

Heidi Andrade, Ed.D., University at Albany—SUNY

Jason Bryer, Ph.D., Excelsior College

Robert Yagelski, Ph.D., University at Albany—SUNY



The contents of this presentation were developed under grant P116F150077 from the U.S. Department of Education. However, those contents do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the Federal Government.



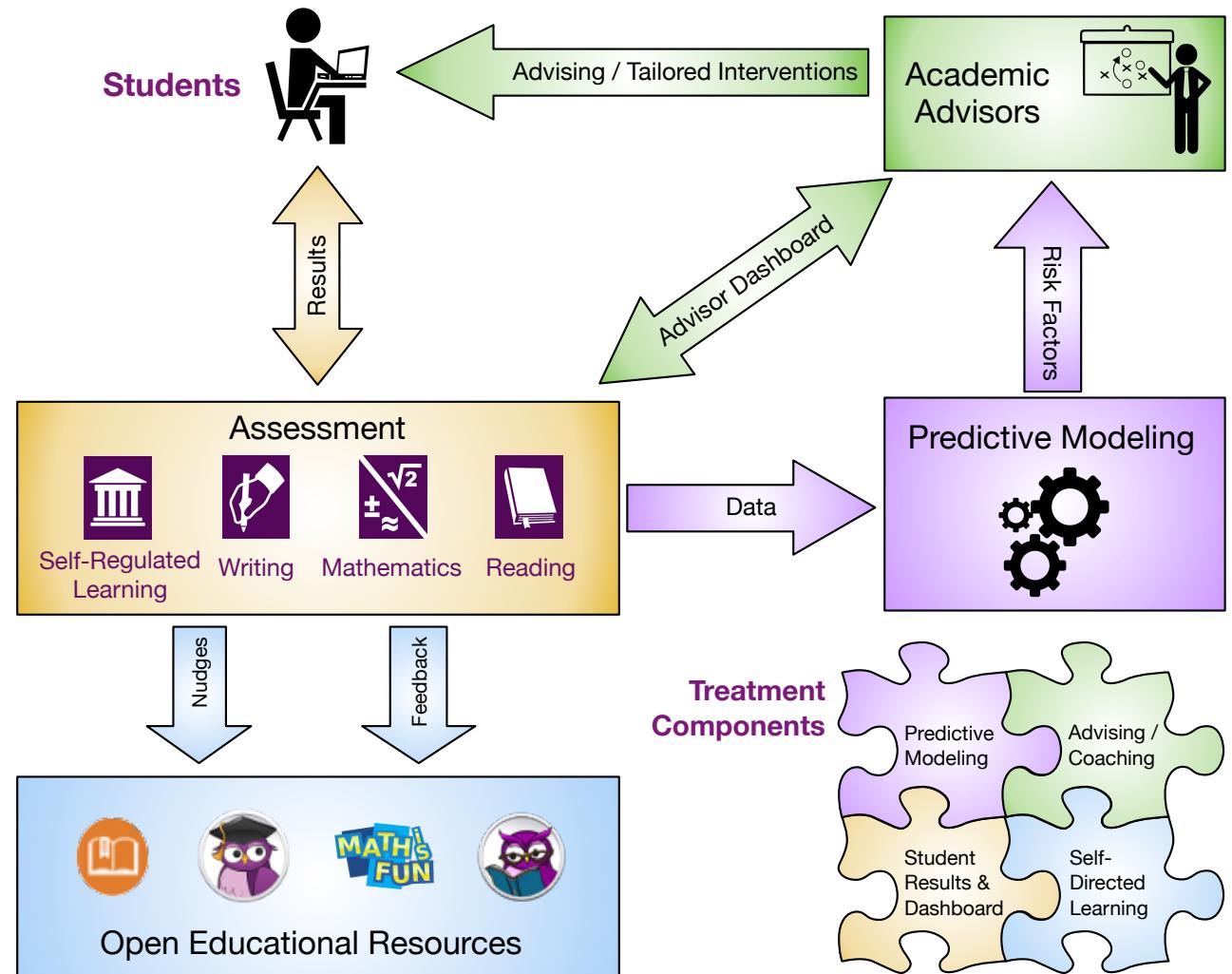
Outline

- What is DAACS?
- The Writing Assessment
- Validity and Reliability of the Writing Assessment
 - Validity
 - Reliability of Human Raters
 - Reliability of LightSide
- Conclusions and Future Directions

What is DAACS?

A suite of diagnostic assessments, technological and social support that work together to optimize learning.

Try: demo.daacs.net



myDAACS - DAACS Jason

Secure | https://demo.daacs.net/dashboard

DAACS
Diagnostic Assessment & Achievement of College Skills

HOME ABOUT MYDAACS LOG OUT

SELF-REGULATED LEARNING WRITING READING MATHEMATICS

COMPLETED COMPLETED COMPLETED COMPLETED

Welcome to the writing assessment! The purposes of this assessment are to:

- provide an opportunity to reflect on your self-regulated learning survey results;
- give you feedback about your writing and suggestions for improving it, as needed;
- direct you to online resources that can help you strengthen your writing skills.

You will be asked to write a one- or two-page reflection on your SRL survey results. After your writing is assessed according to the criteria on the next page, you will receive feedback and suggestions intended to help you become a confident, successful writer in college.



Self-Regulated Learning - DAA x Jason

Secure | https://demo.daacs.net/assessments/college_skills

DAACS
Diagnostic Assessment & Achievement of College Skills

HOME ABOUT MYDAACS LOG OUT

SELF-REGULATED LEARNING

Show Results From March 18, 2018 10:01 AM RETAKE ASSESSMENT

Metacognition Strategies Motivation

myDAACS » Self-Regulated Learning

Overview

Thank you for completing the Self-Regulated Learning (SRL) component of the Diagnostic Assessment and Achievement of College Skills (DAACS). The results presented here are meant to be informative and instructional; **there is no passing or failing**. Because your results can change due to life circumstances or courses you take, you are encouraged to take the SRL assessment again, especially when beginning new courses.

DAACS - SRL from DAACS

DAACS™
Diagnostic Assessment & Achievement of College Skills

Purposes of the Writing Assessment

1. Evaluate incoming students' writing skills
2. Give students targeted, actionable feedback about critical elements of their writing
3. Direct students towards relevant writing resources
4. Assist students in reflecting on their DAACS results and committing to a course of action related to their self-regulated learning
5. Provide supplemental information to academic advisors about students' strengths and weaknesses in terms of SRL.

This writing assessment is not designed to place students, nor to measure their growth in writing ability.

Writing Assessment Prompt



Assessment: **Writing**

Writing

You received information about your learning skills after you took the self-regulated learning (SRL) survey, as well as suggestions for becoming a more effective and efficient learner. Now, in order to reflect on your learning skills and receive feedback on your writing, please use the results from your SRL survey to do your best writing in a brief essay that answers the questions below.

You will need to refer to your SRL survey results and feedback in your essay. We recommend reviewing them, taking notes, and then returning here to write.

Essays must be at least 350 words in order to be meaningfully scored. Please aim to write a complete, well-developed essay in order to get accurate feedback about how ready you are for academic writing, and what you can do to strengthen your writing skills.

- What do your self-regulated learning survey results and the feedback tell you about your learning skills? Use results from the survey and the feedback to support your analysis.
- Which suggested strategies from the feedback are you committed to using this term? Explain why you are committed to using those strategies.

Criteria seen by students

You will receive feedback about your writing according to the following criteria.

Content

- the essay uses the survey results and feedback to create a detailed summary of your strengths and weaknesses as a learner,
- contains suggestions you are committed to using,
- explains your choices of suggestions in terms of your survey results and feedback.

Organization

- the essay has a clear and logical organization,
- uses transitions and linking words and phrases to guide readers through the discussion.

Paragraphs Focus on a Main Idea

- paragraphs consistently and clearly focus on a main idea or point.

Connecting Ideas Within Paragraphs

- sentences are linked together in a way that allows the reader to see the relationship between the ideas or information in one sentence and those in another sentence.
- uses adverbs (e.g., similarly, also, therefore), relative pronouns (e.g., who, that, which), and conjunctions (e.g., and, or, while, whereas) to link sentences and ideas together.

Sentences

- sentences are correct: no run-ons, fragments, or errors in subject-verb agreement.
- uses a variety of sentence structures.

Conventions

- spelling, punctuation, and capitalization are correct.

Scores and feedback seen by students

The screenshot shows a web browser window titled "Structure - Organization - Writing" with the URL <https://demo.daacs.net/assessments/writing/domain/organization/subdomain/structure?takenDate=2018-03-18T12%3A08%3A21.480Z>. The user is logged in as "Jason".

STRUCTURE

Show Results From: March 18, 2018 8:08 AM

RETAKE ASSESSMENT

	Mastering	Emerging	Developing
Structure	The essay is well-organized, with an order and structure that present the discussion in a clear, logical manner.	The essay has a basic structure but does not offer a clear, overall organization that enables a reader to understand the progression of one idea to another. It might have been strengthened by the inclusion of an introduction, topic sentences, conclusion, etc.	The essay is not carefully structured. It might move almost randomly from one point to the next. It could have been strengthened by the use of an introduction, topic sentences, a conclusion, and other organizational features.

My scores indicate that I need to start evaluating my work, whether that be during or after the process of completion. Will I make an effort to change? Yes, to an extent. I say that because I am not one to doubt my work. My confidence level on what I turn in, is always one hundred percent. That's because I always make sure that whatever it is I'm turning in, is the best version it can be. That being said, for me to doubt my work during and after the process that's like me doubting my credibility and ability to finish something. I then would be putting setbacks on my performance. I'm all for learning and improving, but I will never doubt my work. I might evaluate it more during the process... [SEE MORE](#)

Additional Resources

Scoring Rubric Used by Raters

Criteria	Developing (1)	Emerging (2)	Mastering (3)
Content	Summary The discussion of the survey and feedback is vague, poorly grounded in the survey results and feedback, and/or simplistic.	The essay uses evidence from survey results and feedback to summarize student's strengths and weaknesses in terms of self-regulated learning. The summary lacks sufficient detail; might be under-developed in places, e.g., strengths or weaknesses might get short shrift.	The essay uses relevant survey results and feedback to provide a detailed summary of both the student's strengths <u>and</u> weaknesses in terms of self-regulated learning.
	Suggestions Choices of suggestions to which to commit are vague, if present at all, and/or only loosely connected to the survey results and feedback, if at all. The essay might refer to the continued use of current strategies but not to anything new related to the SRL feedback.	Choices of suggestions to which to commit are discussed. The connections to the survey and feedback are present but might not always be explicit.	The discussion of suggestions for improvement in SRL are logically and explicitly related to the survey results and feedback, and developed in sufficient depth.
Organization	Structure The structure and order of the essay is weak, unclear, and/or illogical.	The essay has a general structure and order but may not have a clear overall organization that enables a reader to follow the progression of one idea to another. Although the structure is logical, it might seem haphazard at times. Note: One-sentence paragraphs do not necessarily reflect a problem with organization, but numerous such paragraphs might signal a weak or haphazard structure.	The essay is well-organized, with an order and structure that present the discussion in a clear, logical manner.
	Transitions Transitions between paragraphs are missing or ineffective; paragraphs tend to abruptly shift from one idea to the next. Note: One-paragraph essays receive a 1 for this criterion.	Paragraphs are usually linked with transitions, as needed. The transitions might be implied or strained, but the reader can follow along.	Transitions between paragraphs are appropriate and effective, and strengthen the progression of the essay (e.g. "The second aspect . . ." "The last aspect . . ." and/or the repetition of important ideas and terms to connect paragraphs).
Paragraphs	Focus on a Main Idea Most or all paragraphs lack one clear, main point; might have several topics. Note: Numerous brief paragraphs of one or two sentences each might indicate a problem with paragraph focus and warrant a score of 1.	Paragraphs are generally but not consistently focused on a main idea or point. Some paragraphs might lack a clear focus in an essay in which the majority of paragraphs maintain a clear focus on a main idea.	Paragraphs are consistently and clearly focused on a main idea or point.
	Cohesion The connections between ideas in sentences within paragraphs are unclear. Little effective use of linking words and phrases.	The ideas or information in each sentence within a paragraph are generally but not consistently linked together, if only loosely. Additional or better choices of linking words and phrases would clarify the connections b/w ideas within paragraphs.	Within paragraphs, the individual sentences are seamlessly linked together; the reader can see the relationship between the ideas or information in one sentence and those in another sentence. The writing explicitly links sentences and ideas using adverbs (e.g., similarly, also, therefore), relative pronouns (e.g., who, that, which), conjunctions (e.g., and, or, while, whereas), and/or the repetition of key words, as appropriate.
Sentences	Correct Significant syntax problems, such as fragments, run-on sentences, missing/extraneous words, awkward constructions, dangling modifiers, and/or transposed words, are present and numerous enough to distract readers and impede meaning.	Grammatically incorrect sentences, when present, are minor and do not interfere with meaning.	There are very few or no significant syntax problems. The writer is capable of managing even complex syntactic structures correctly.
	Complex The sentences lack syntactic complexity and vary little, if at all, in structure. The sentences tend to be relatively simple in structure, following a basic subject-verb-object pattern perhaps with a few additional elements, such as brief introductory phrases, prepositional phrases, or modifiers.	Complex syntactic structures are present but may not always be managed effectively; sentence structures may be varied but are not often sophisticated.	Consistent and appropriate use of a variety of sentence structures, including sophisticated sentence structures, such as complex, compound, or compound-complex sentences, and other complex syntactic forms, such as extended participial phrases and relative clauses.
Conventions	A pattern of errors in spelling, punctuation, usage (such as incorrect word forms or subject-verb agreement), and/or capitalization suggest that the writer struggles with the rules for conventions.	Spelling, punctuation, usage, and capitalization are generally correct. There may be errors but there is no pattern that suggests that the writer struggles with the basic rules.	Spelling, punctuation, and capitalization are correct to the extent that almost no editing is needed. There are very few, if any, very minor errors of usage.

Validity

- *Validity*: “The degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (AERA, APA, & NCME, 2014, p. 11)
- Process of validation involves collecting *different sources of evidence* to provide scientific basis for the proposed score interpretations
 - **Test content**: alignment between test content, content domain, and proposed interpretations of test scores
 - Response processes
 - **Internal structure**: degree to which relationships among test items and components conform to conceptualization of measured construct.
 - Relations to other variables
 - Testing consequences

Validity: Test Content

Phase 1: Development of the writing assessment prompt and rubric

Phase 2: Iterative cycles of reviews and revisions

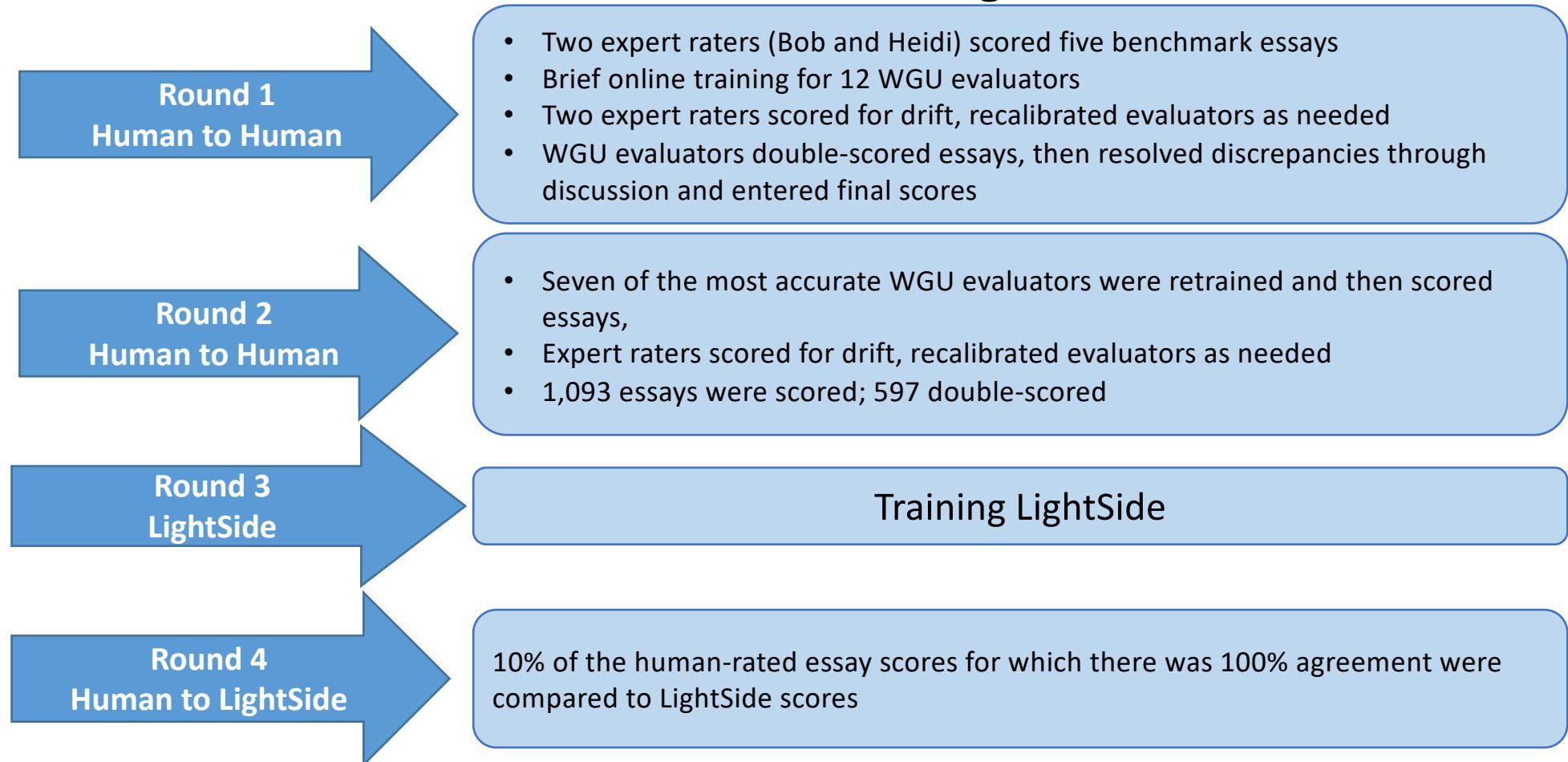
Phase 3: Pre-pilot of the writing assessment

Phase 4: Revisions to writing assessment feedback (in progress)

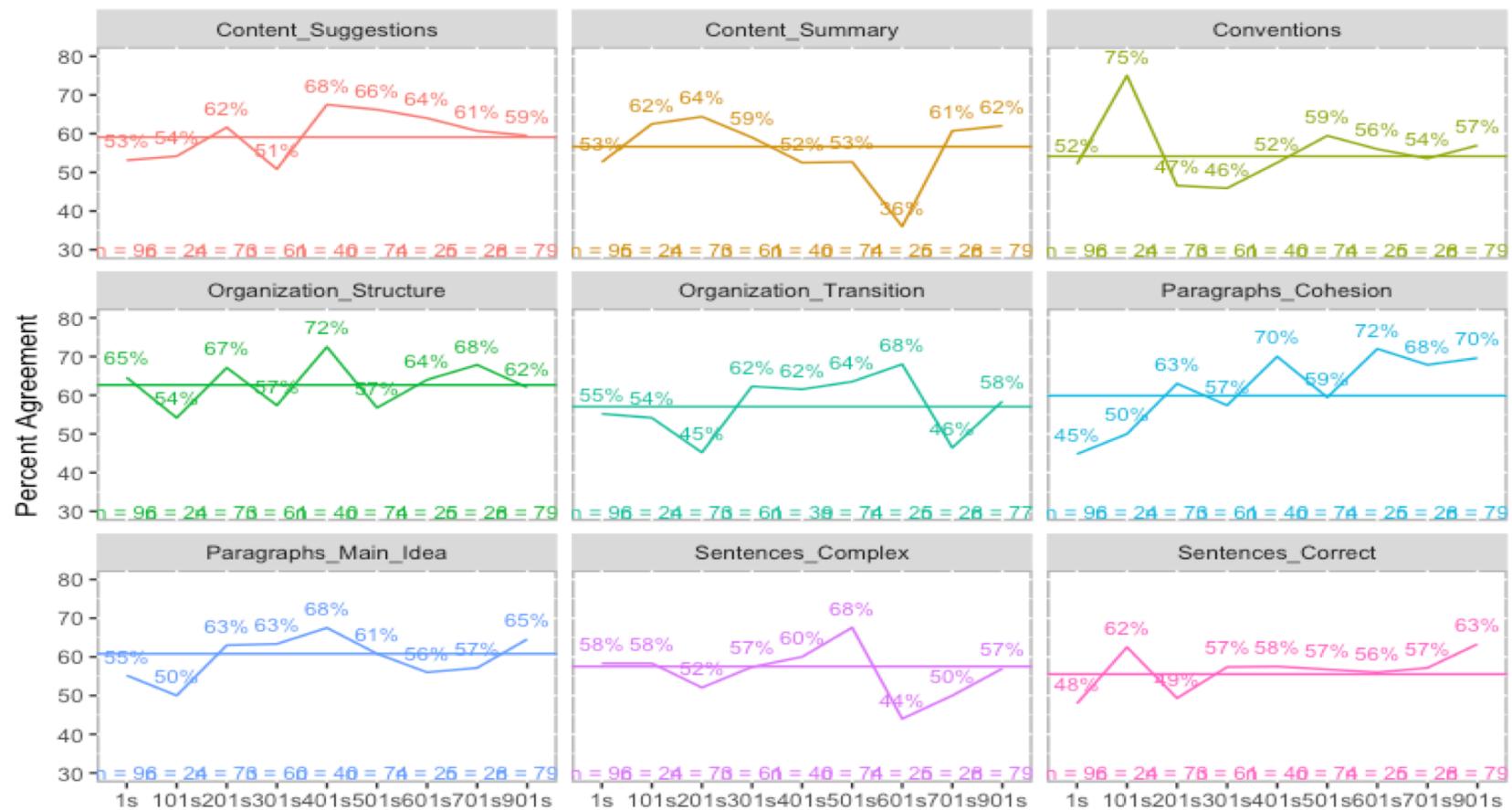
Reliability: Procedures

- Score and collect evidence of **reliability** b/w human scorers, and b/w humans and LightSide
 - Phase One: Human to human
 - Two expert raters (Bob and Heidi) scored five benchmark essays
 - Brief online training for 12 WGU evaluators
 - Two expert raters scored for drift, recalibrated evaluators as needed
 - WGU evaluators double-scored essays, then resolved discrepancies through discussion and entered final scores
 - Phase Two: Human to human
 - Seven of the most accurate WGU evaluators were retrained and then scored essays
 - Expert raters scored for drift, recalibrated evaluators as needed
 - Total of 1,093 essays were scored; 597 double-scored in phases one and two
 - Phase Three: Train LightSide
 - For each criteria, only the essays where the two human raters initially agreed were included in the training dataset (n ranged from 247 to 287, x = 266). *As a result, the training datasets had 100% human-to-human agreement.*
 - 10-fold validation was used (i.e. models were trained using 90% of the essays, 10% was withheld for validation)
 - Phase Four: Comparisons of human to LightSide scores

Reliability: Procedures to score and collect evidence of reliability b/w human scorers, and b/w humans and LightSide



Round 2: Inter-rater reliability for WGU evaluators



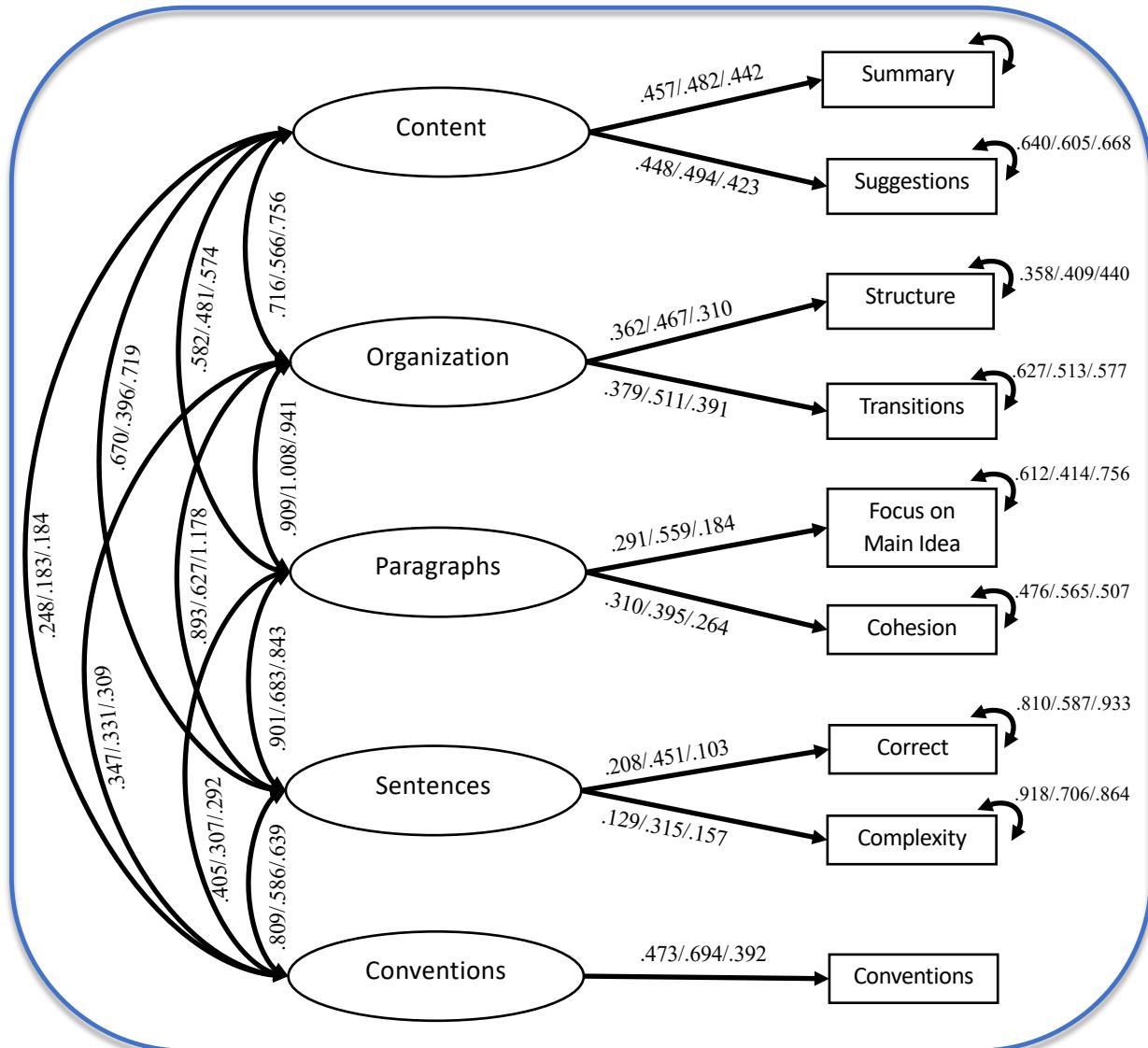
Rounds 3 & 4: Percent Agreement and LightSide Reliability Indices for the first 1000 essays

Criteria	Human to Human % Agreement	Model	Accuracy	Kappa	Bad Error
Summary	55.78 (8.69)	Logit	69.62	0.4179	0.05
Suggestions	59.78 (6.00)	Logit	72.26	0.4912	0.06
Structure	62.89 (5.93)	Bayes	74.22	0.3413	0.01
Transitions	57.11 (7.90)	Bayes	47.17	0.1483	0.04
Ideas	59.78 (5.67)	Logit	73.45	0.1276	0.12
Cohesion	62.67 (9.80)	Logit	72.73	0.2451	0.01
Correct	56.33 (5.05)	Logit	55.73	0.073	0.07
Complexity	56.00 (6.76)	Bayes	68.42	0.361	0.003
Conventions	55.33 (8.54)	Logit	63.16	0.159	0.04

Validity: Internal Structure

Fit Indices of CFA model using Total, Human-Scored, and LightSide-scored Sample

Goodness-of-Fit Indices	Total Sample (n=6618)	Human-Scored (n=879)	LightSide-scored (n=5739)
χ^2 (p-value)	1073.160 (p < .001)	85.264 (p < .001)	840.967 (p < .001)
df	18	18	18
χ^2/df	59.62	4.74	46.72
AIC	80529.651	14648.555	60484.348
CFI	.918	.966	.913
RMSEA	.094	.065	.089
[90% CI]	[.089, .099]	[.052, .079]	[.084, .094]
SRMR	.049	.036	.043



Validity: Internal Structure

Internal Consistency
Reliability Estimates
($n = 6618$)

	Total Sample		Human-Scored		LightSide scored	
	M (SD)	α	M (SD)	α	M (SD)	α
Conventions	2.8 (.47)		2.3 (.69)		2.8 (.39)	
Content	2.5 (.61)	.56	2.3 (.65)	.56	2.6 (.59)	.54
Summary	2.6 (.71)		2.3 (.78)		2.6 (.69)	
Suggestions	2.5 (.75)		2.3 (.78)		2.5 (.73)	
Organization	2.5 (.47)	.64	2.3 (.59)	.69	2.5 (.44)	.62
Structure	2.7 (.45)		2.5 (.61)		2.8 (.41)	
Transitions	2.2 (.62)		2.1 (.73)		2.2 (.60)	
Paragraphs	2.8 (.38)	.62	2.5 (.58)	.66	2.9 (.31)	.52
Focus on a Main Idea	2.8 (.47)		2.4 (.73)		2.9 (.37)	
Cohesion	2.8 (.43)		2.5 (.60)		2.8 (.38)	
Sentences	2.5 (.35)	.22	2.4 (.53)	.51	2.5 (.30)	.17
Correct	2.7 (.48)		2.3 (.70)		2.8 (.40)	
Complexity	2.3 (.45)		2.4 (.58)		2.2 (.43)	
TOTAL	2.6 (.32)	.76	2.4 (.42)	.78	2.6 (.28)	.72

Increasing the accuracy of LightSide (LS)

- Experiment #1 regarding Transitions
 - Very little improvement after adding regular expressions, e.g., Also, However, Another, Furthermore, Another area....

Act/Pred	Original (n=265)			Experiment 1.1 (n=265)			
	1	2	3	Act/Pred	1	2	3
1	14	37	7	1	14	37	7
2	12	68	31	2	12	68	31
3	3	50	43	3	4	49	43
Accuracy	47.17			Accuracy	47.17		
Kappa	.1483			Kappa	.1494		
Bad Error	0.0377			Bad Error	0.042		

- Rescoring just two essays on which LS was off by one level resulted in slight improvements:

Act/Pred	Original (n=265)			Experiment 1.2 (n=265)			
	1	2	3	Act/Pred	1	2	3
1	14	37	7	1	16	38	3
2	12	68	31	2	14	71	28
3	3	50	43	3	2	51	42
Accuracy	47.17			Accuracy	48.68		
Kappa	.1483			Kappa	.1693		
Bad Error	0.0377			Bad Error	0.018		

Increasing the accuracy of LightSide (LS)

- Experiment #2: Adding scores for 5 additional essays
 - Mostly slight improvements in terms of individual criteria
 - Take, for example, the results for the Correct Sentences criterion:

Original (n=253)				Experiment 2.1 (n=258)			
Act/Pred	1	2	3	Act/Pred	1	2	3
1	0	5	18	1	0	5	18
2	0	20	68	2	0	21	68
3	0	21	121	3	0	21	125
Accuracy	55.73			Accuracy	56.59		
Kappa	0.073			Kappa	.0848		
Bad Error	0.071			Bad Error	0.070		

Increasing the accuracy of LightSide (LS)

- Experiment #3: LS was producing total scores that, on average, were higher than WGU scores on a different sample. We scored five of the lowest-scoring essays. Of 19 score discrepancies b/w us and LS, 14 were scored lower by us, suggesting that LS is scoring too leniently.
 - Slight improvements on some criteria when the 5 lowest-scoring essays were added to the training sets, e.g., the Structure criterion:

Original (n=287)				Experiment 4 (n=292)			
Act/Pred	1	2	3	Act/Pred	1	2	3
1	0	1	4	1	0	3	4
2	0	37	44	2	0	43	41
3	0	25	176	3	0	26	175
Accuracy	74.22			Accuracy	74.66		
Kappa	0.3413			Kappa	.3826		
Bad Error	0.01			Bad Error	.01		

Conclusions and Future Directions

- Scale up experiments to increase accuracy of LightSide
- Address issues with measuring and interpreting IRR metrics
 - ICC versus percent rater agreement

Thank You!

www.DAACS.net

Heidi Andrade, Ed.D., handrade@albany.edu

Jason Bryer, Ph.D., jbryer@excelsior.edu

Robert Yagelski, Ph.D., ryagelski@albany.edu