Walter: Transforming Assessment with GenAI

- Insights from Groundbreaking Comparative Studies

David M DiSabito Jr, MBA
Sanjeev Jha, PhD
Manshika Chakravarthy Nalla
Supreeth Reddy Mamilla
Sri Vastava Reddy Banda

with special thanks to:
Lisa Hansen, Thomas Mennella, and Josephine Rodriguez

Prepared for Presentation at

AMCOA

April 25, 2025

WESTERN NEW ENGLAND UNIVERSITY WNE

WNE: Who Are We?

- Private, doctoral/professional University in Springfield, MA
- 2584 undergraduates & 990 graduate students
- 5 Academic Units:
  - College of Arts and Sciences
  - College of Business
  - College of Engineering
  - College of Pharmacy and Health Sciences
  - School of Law

Goals of this Presentation

- Address concerns about using GenAI in institutional and course-level assessment.
- Introduce Walter, a GenAI app transforming assessment and accreditation.
- Show Walter as a second reader for goals, objectives, evidence, and rubrics.
- Present case study comparing human vs. AI scoring with identical rubrics.

…Goals of this Presentation

- Highlight Walter's speed in evaluating large volumes of student work.
- Ensure Walter's privacy for FERPA compliance.
- Explore benefits, challenges, and ethics of AI in assessment.
- Discuss Walter's implementation and impact on assessment strategies.

Overview of Institutional Assessment

Best Practices

- Authentic Assessments
- Aligned with LO's
- Clearly Defined Rubrics
- Training & Norming
- Continuous Improvement
- Meaningful, Measurable & Manageable

Common Challenges

- Data Collection & Analysis
- Resource Constraints
- Unconscious Bias
- Academic Complexity
- Engaging Faculty
- Sustaining Commitment

Potential Role of GenAI

Consistently and efficiently applies grading criteria across all student work

Promotes an objective, standardized, transparent assessment

Does not get tired or experience fatigue

Produces immediate formative feedback for students

Mitigates unconscious human bias & errors (…?)

AI could help humans foster a more efficient and objective assessment environment.

Potential Bias in Assessment

Traditional Assessments

- **Instructor-Student Relationship** (lenience, strictness)
- **Implicit Biases** (Race, gender, socioeconomic status, culture,…)
- **Grading Inconsistencies** (Fatigue, mood, distractions,…)

AI-Assisted Assessment

- **Inherent Bias** (Gen AI inherits societal biases of training data)
- **Flaws in Sampling** (underrepresented populations in training data)
- **Predictive Text Bias** (Echo chamber of public domain)

Motivation for WNE Research Study

- Can GenAI be used to score work using a rubric?
- Can the assessment be done in a way that seems "reasonable" to an instructor?
- Can the drudgery of assessment be reduced?
- Can faculty then spend their time **discussing the results and planning for improvements in teaching and learning**?

Ethical Implications

- **Data Privacy** - Privacy concerns arise when using student data/evidence with GenAI models
- **Transparency** – Educators need to be open with students, colleagues, and administrators when/if they use GenAI for assessment purposes
- **Student Consent** – Essential to get informed consent from students when their work will be assessed by GenAI

Walter – Transforming Assessment, Empowering Educators!

- **Web Based Cloud Application**
- **Detailed Reporting**
  Instantaneous Results
  Improved Feedback Cycle
- **More Time to focus on Strategic Priorities**
  More Efficiency and Less Drudgery
- **Privacy – Zero Data Retention Policy**
  Business Associate Agreement
  AI Integrated Concepts, Inc. and OpenAI, L.L.C.
- **Speed and Accuracy**
  Socially Responsible Business Practices
  Learning Goal 4 Learning Objective 1 (LG4LO1)
  55 documents in 25 seconds.

WNE CoB BSBA
Learning Goal 4 Learning Objective 1 (LG4LO1)

SCORING RUBRIC – BSBA LG 4

LEARNING GOAL 4:	Demonstrate understanding of socially responsible business practices.
Learning Objective 1:	Identifies different strategies for an organization to demonstrate socially responsible business practices. - Three-point scoring scale. 3: Exceeds 2: Meets 1: Fails

College of Business
WESTERN NEW ENGLAND UNIVERSITY | WNE

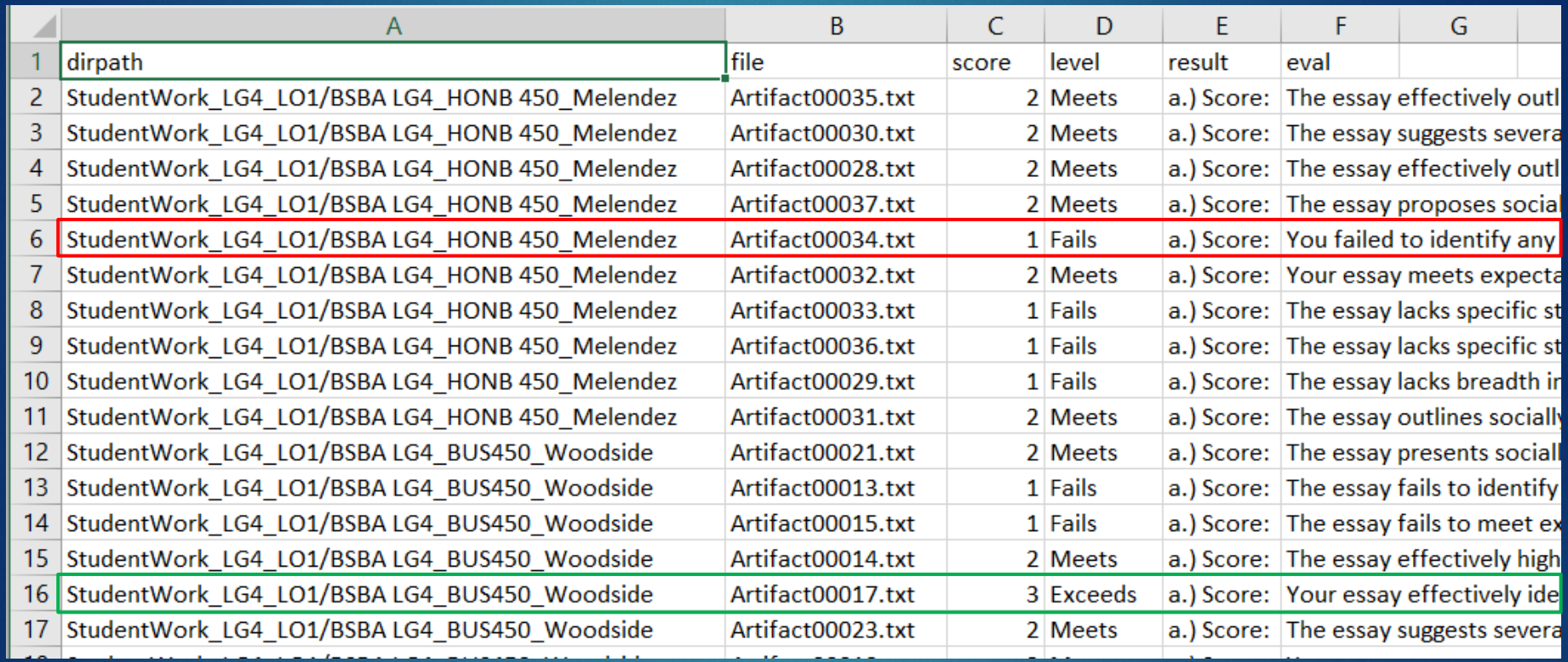
AACSB ACCREDITED

Demonstration Overview

- Logon
- Projects
- Project Parameters
- Copy Project
- Edit Project
- Download Results
- Run Project
- New Project
  - Title
  - User Instructions/Sample Output
  - Rubric
  - Discipline
  - Critique Length
  - GenAI - Source, Model
  - Grading Conversion

Reports\Exports in .csv format

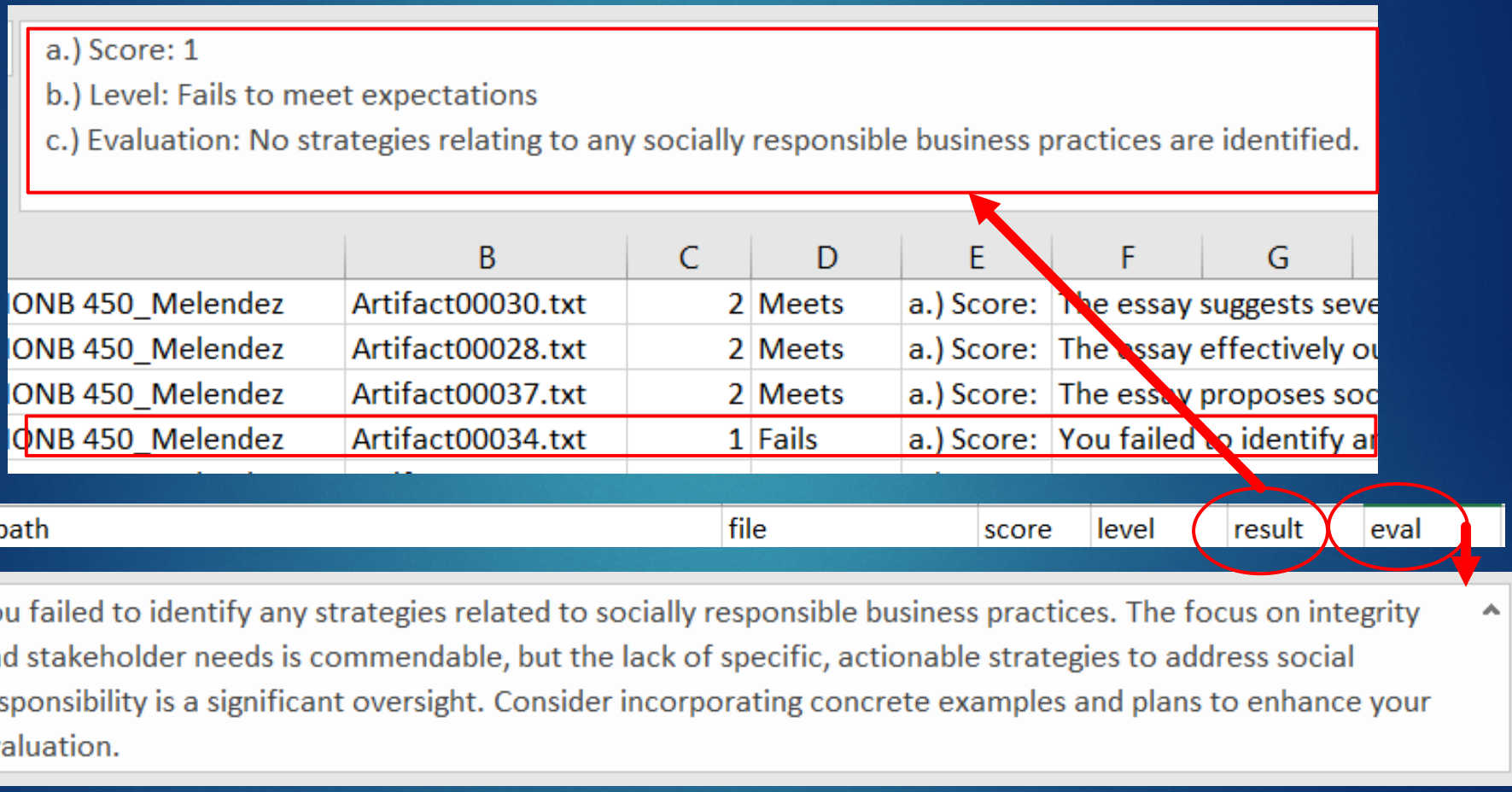
Easy to read in MS Excel or Google Sheets.



	A	B	C	D	E	F	G
1	dirpath	file	score	level	result	eval	
2	StudentWork_LG4_LO1/BSBA LG4_HONB 450_Melendez	Artifact00035.txt	2	Meets	a.) Score:	The essay effectively outl	
3	StudentWork_LG4_LO1/BSBA LG4_HONB 450_Melendez	Artifact00030.txt	2	Meets	a.) Score:	The essay suggests severa	
4	StudentWork_LG4_LO1/BSBA LG4_HONB 450_Melendez	Artifact00028.txt	2	Meets	a.) Score:	The essay effectively outl	
5	StudentWork_LG4_LO1/BSBA LG4_HONB 450_Melendez	Artifact00037.txt	2	Meets	a.) Score:	The essay proposes social	
6	StudentWork_LG4_LO1/BSBA LG4_HONB 450_Melendez	Artifact00034.txt	1	Fails	a.) Score:	You failed to identify any	
7	StudentWork_LG4_LO1/BSBA LG4_HONB 450_Melendez	Artifact00032.txt	2	Meets	a.) Score:	Your essay meets expecta	
8	StudentWork_LG4_LO1/BSBA LG4_HONB 450_Melendez	Artifact00033.txt	1	Fails	a.) Score:	The essay lacks specific st	
9	StudentWork_LG4_LO1/BSBA LG4_HONB 450_Melendez	Artifact00036.txt	1	Fails	a.) Score:	The essay lacks specific st	
10	StudentWork_LG4_LO1/BSBA LG4_HONB 450_Melendez	Artifact00029.txt	1	Fails	a.) Score:	The essay lacks breadth in	
11	StudentWork_LG4_LO1/BSBA LG4_HONB 450_Melendez	Artifact00031.txt	2	Meets	a.) Score:	The essay outlines socially	
12	StudentWork_LG4_LO1/BSBA LG4_BUS450_Woodside	Artifact00021.txt	2	Meets	a.) Score:	The essay presents social	
13	StudentWork_LG4_LO1/BSBA LG4_BUS450_Woodside	Artifact00013.txt	1	Fails	a.) Score:	The essay fails to identify	
14	StudentWork_LG4_LO1/BSBA LG4_BUS450_Woodside	Artifact00015.txt	1	Fails	a.) Score:	The essay fails to meet ex	
15	StudentWork_LG4_LO1/BSBA LG4_BUS450_Woodside	Artifact00014.txt	2	Meets	a.) Score:	The essay effectively high	
16	StudentWork_LG4_LO1/BSBA LG4_BUS450_Woodside	Artifact00017.txt	3	Exceeds	a.) Score:	Your essay effectively ide	
17	StudentWork_LG4_LO1/BSBA LG4_BUS450_Woodside	Artifact00023.txt	2	Meets	a.) Score:	The essay suggests severa	

Reports\Exports in .csv format

Easy to read in MS Excel or Google Sheets.



The next case study is **groundbreaking** for two reasons:

1.) The **students, human instructor, and AI all used the same rubric**.

2.) Rather than just comparing overall human and AI score totals, we further **evaluated the scoring comparisons from the rubric's criteria level**.

WNE Case Studies

We wanted to determine if humans and AI assess student evidence the same.

Our null hypothesis assumes they do.

Our alternative hypothesis is that they do not.

We used a matched pair *t*-test and the correlation coefficient to analyze the results.

WNE Case Studies

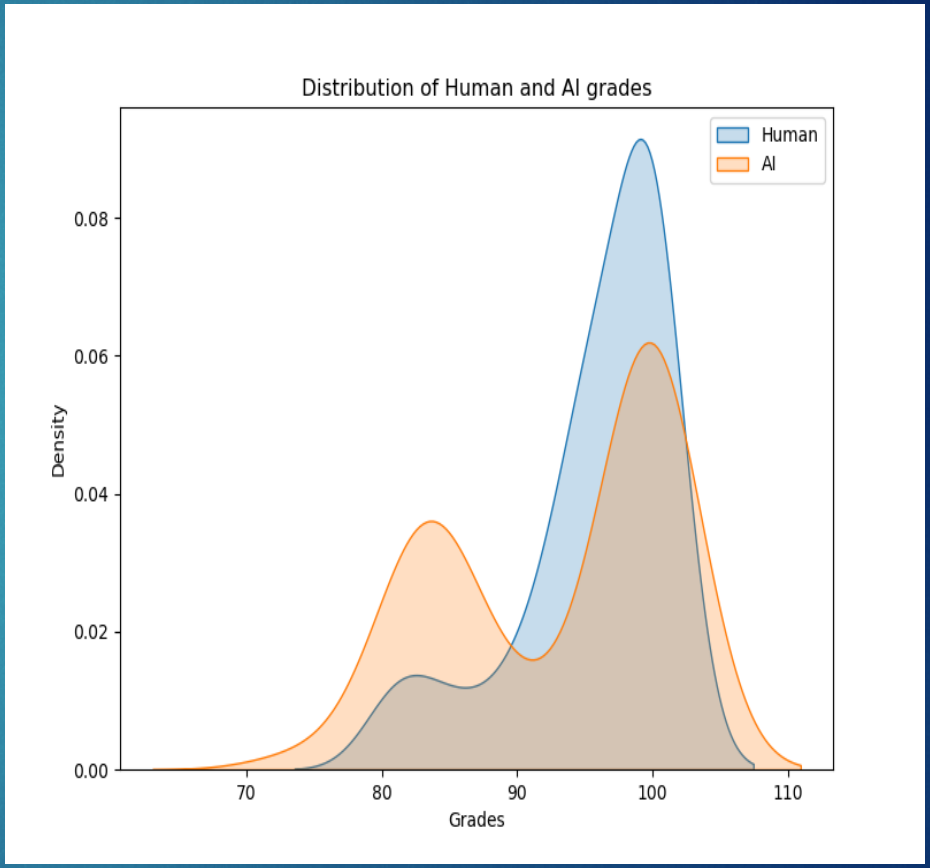
- Assessment Types:
  - Course-level (Instructor scoring) - BAIM 202 - Writing
  - Institutional Assessment (Team based scoring) - LG4LO1 - Writing

Case Study 1: BAIM 202

100 Point Scale

Business Information Systems Paper

Sample size:	57
Human mean:	95.74
AI mean:	93.11
Alpha:	0.05
T-statistic:	2.20
P-value:	.0318
Correlation:	.187



Significant Difference in Means

Low Correlation

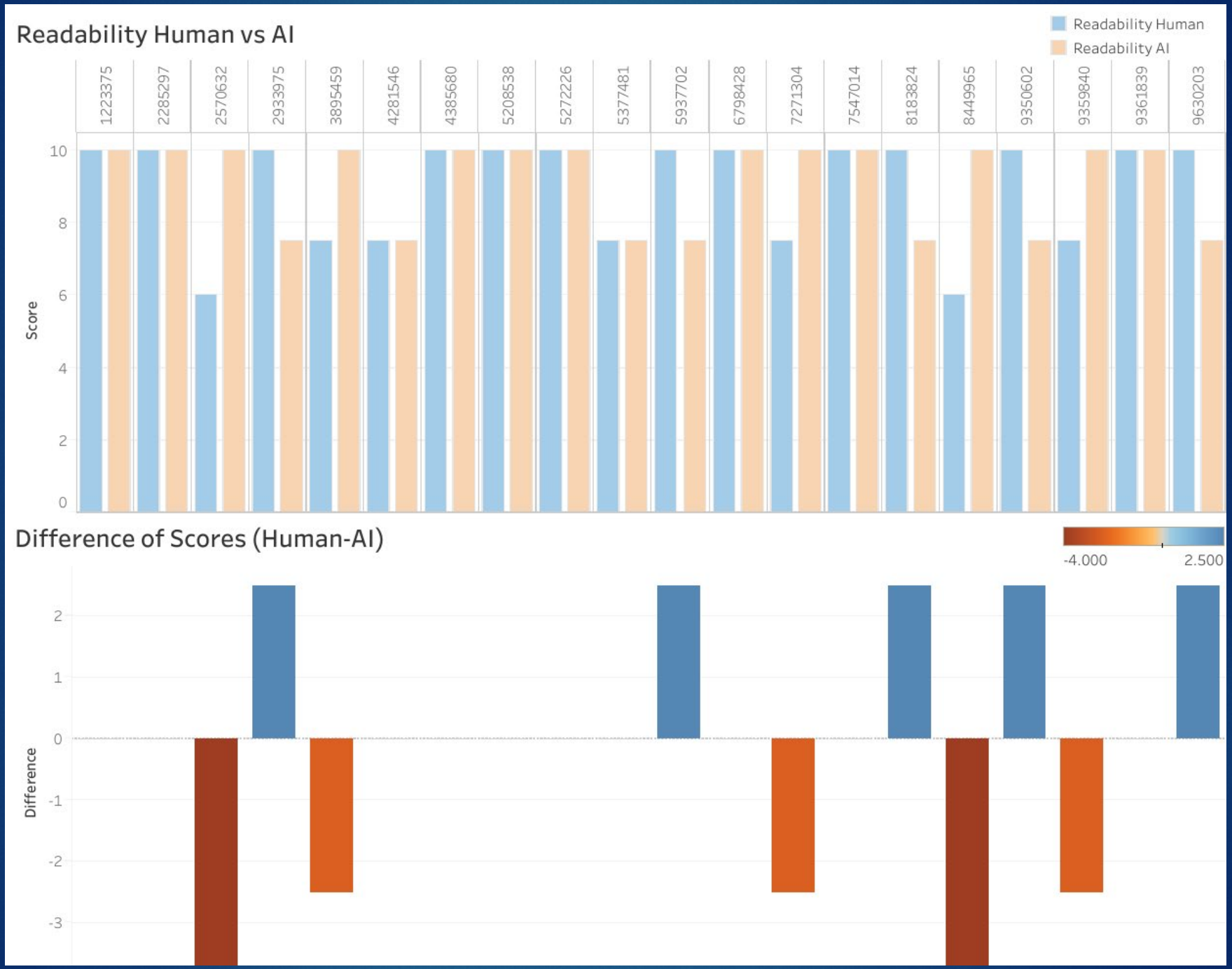
Rubric Criteria used-

- **Readability**
- Business Process Discussion
- IS Discussion
- Information Discussion
- Information Qual Discussion
- **Business IS Information**

Criteria 1- Readability
10 points

- Rubrics used for Human and AI

Level	Points	Description
Excellent	10	No grammatical errors; all sections (Intro, Body, Summary, Citations) present; correct citation format.
Fair	7.5	1-3 grammatical errors; all sections present but may lack clarity; mostly consistent citations.
Poor	6	4-6 grammatical errors; one section missing or incomplete; inconsistent citation format.
Fail	5	More than 6 grammatical errors; two or more sections missing or incomplete; inconsistent citation format.
Zero	0	Paper is missing or no recognizable sections.


Readability Human vs AI
Readability Human
Readability AI
1223375
2285297
2570632
2933975
3895459
4281546
4385680
5208538
5272226
5377481
5937702
6798428
7271304
7547014
8183824
8449965
9350602
9359840
9361839
9630203
Score
10
8
6
4
2
0
Difference of Scores (Human-AI)
-4.000
2.500
Difference
2
1
0
-1
-2
-3

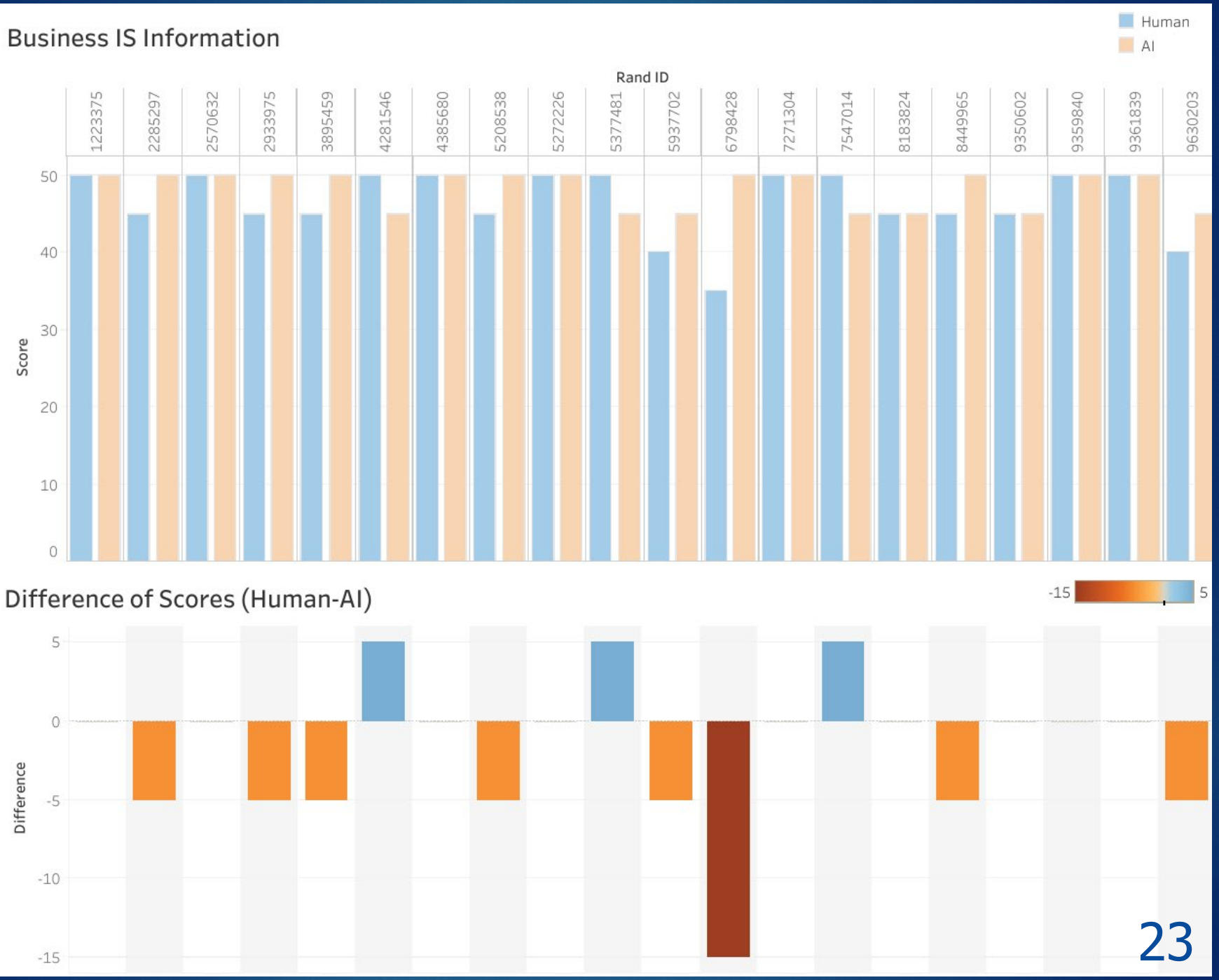
Criteria 6- Business IS Information

50 Points

- Rubrics used for Human and AI

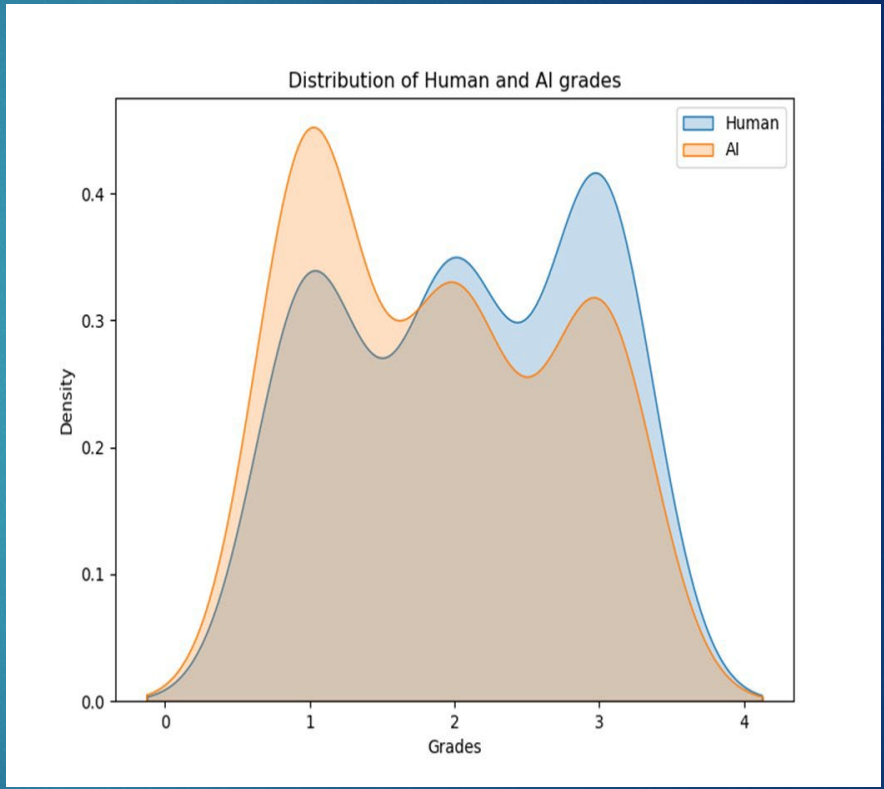
RELATIONSHIP BETWEEN BUSINESS PROCESS, INFORMATION SYSTEMS, AND INFORMATION

Level	Points	Description
Excellent	50	Comprehensive and clear explanation of the interrelation among all three components.
Fair	45	Adequate explanation, covers most interrelations but lacks detail or clarity in one or two areas.
Poor	40	Limited explanation; covers some interrelations with minimal detail and clarity.
Fail	35	Inadequate explanation; very few interrelations are covered or explained.
Zero	0	No explanation or evidence of understanding the interrelation among the components.


Business IS Information
Human
AI
Rand ID
1223375
2285297
2570632
2933975
3895459
4281546
4385680
5208538
5272226
5377481
5937702
6798428
7271304
7547014
8183824
8449965
9350602
9359840
9361839
9630203
Score
50
40
30
20
10
0
Difference of Scores (Human-AI)
-15
5
Difference
5
0
-5
-10
-15

Case Study 2: LG4_LO1- Socially Responsible Business Practices

Socially Responsible Business Practices (3 pts.)	
Sample size:	55
Human mean:	2.07
AI mean:	1.87
t-statistic:	2.11
p-value:	.0399
Correlation:	.647

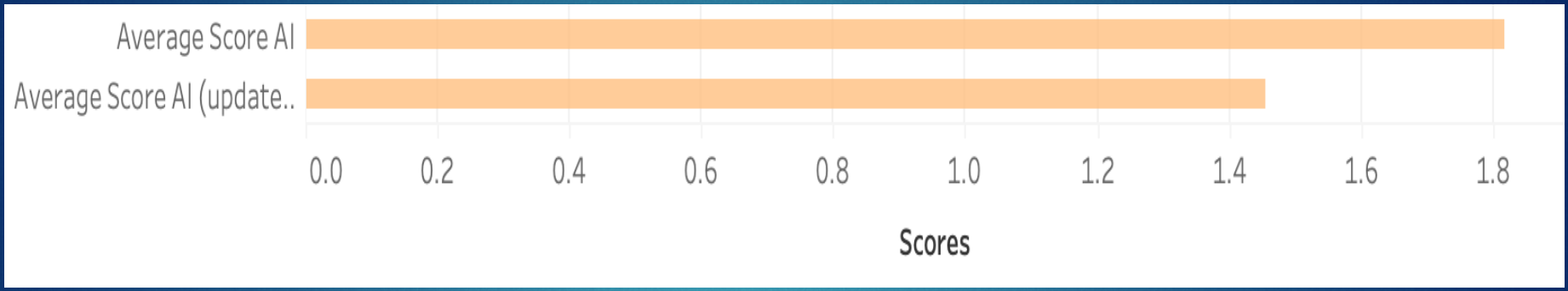


Significant Difference in Means
Moderately High Correlation

Original Rubric Used By Human and AI

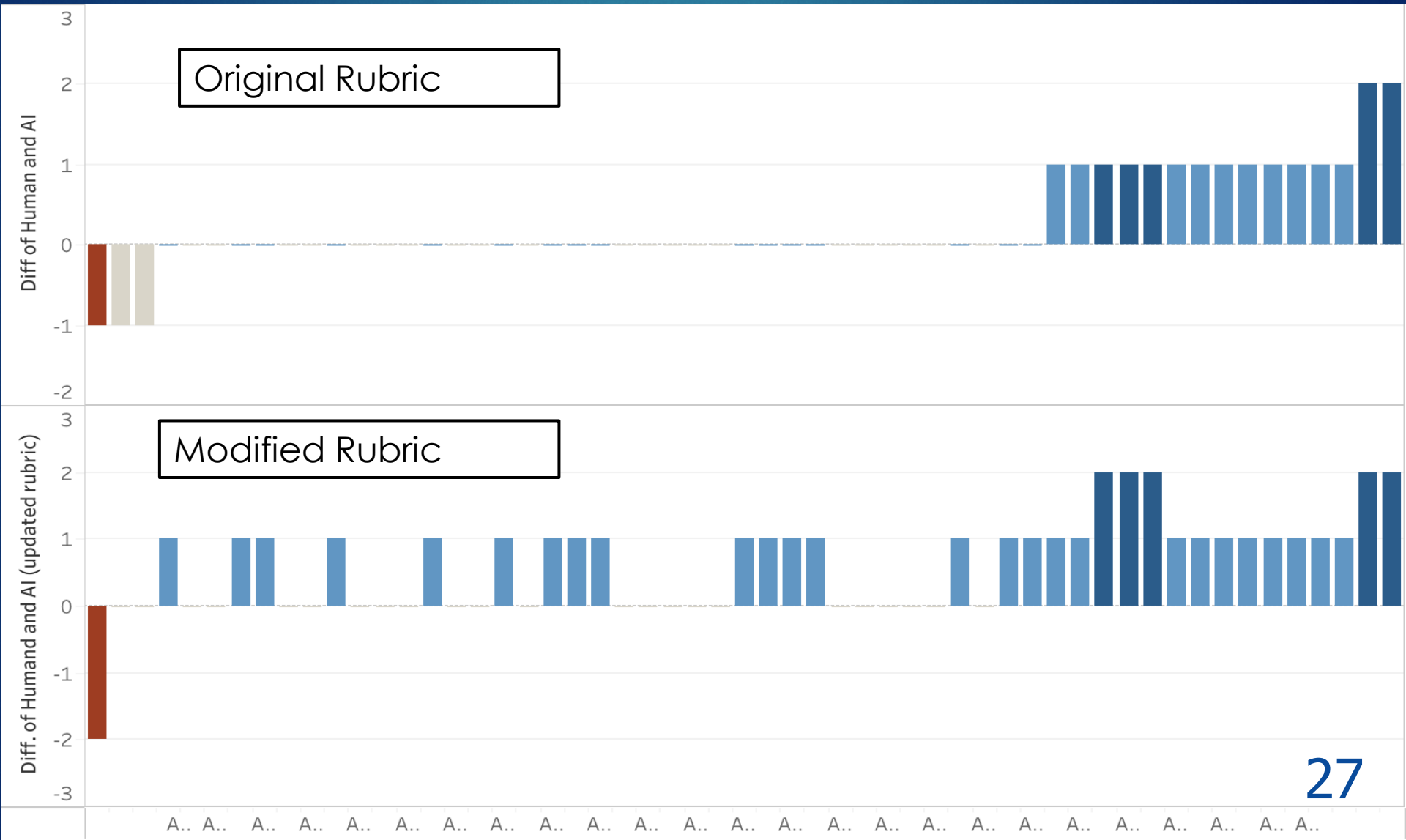
Level	Points	Description
Exceeds expectations	3	There are **multiple** socially responsible business practices identified that are applicable in the real world
Meets expectations	2	The writer **suggests** socially responsible strategies but these may not be pragmatic.
Fails to meet expectations	1	**No** strategies relating to any socially responsible business practices are identified.

AI score for a modified rubric



Level	Points	Description
Exceeds expectations	3	There are **at least 3** socially responsible business practices identified that are applicable in the real world **with examples**
Meets expectations	2	The writer suggests **at least 2** socially responsible strategies but these may not be pragmatic.
Fails to meet expectations	1	**Only 1 or no** strategy relating to any socially responsible business practices is identified.

Difference in Human and AI Scores



Insights and Takeaways

Potential for AI to handle more routine assessment tasks and provide faculty with more time to spend on higher order aspects

AI may be able to reduce institutional assessment cycle times

Provides quick opportunity to clarify rubrics which can improve teaching and learning

Human assessment may still provide instructors with deeper understanding of student learning

Insights and Takeaways

Significant upfront time and resources to develop AI tools

Does not perfectly replicate human judgment

Struggles with handwritten input and distinguishing sources

Requires precise rubrics and instructions

Loses the "human touch" of assessment

Our Next Steps

Assurance of Learning (AoL) with AACSB Enhancement Proposal

Incorporating AI as a 2nd Reader

College of Business,

Western New England University

Objective:

- Enhance assessment methodology by leveraging human evaluators and AI tools
- Improve objectivity, equity, and efficiency in student assessments

Question to Consider:

Are humans the…

?

Thank You!

Contact us:

David M DiSabito Jr, MBA
AI Liaison to College of Business
Professional Educator of Business Analytics and Information Management
cell 413.348.1963
david.disabito@wne.edu

Sanjeev Jha, PhD
AACSB Coordinator
Associate Professor of Business Analytics and Information Management
sanjeev.jha@wne.edu

Manshika Chakravarthy Nalla
Teaching Fellow
Graduate Research Assistant
manshikachakravarthy.nalla@wne.edu

Supreeth Reddy Mamilla
Teaching Fellow
Graduate Research Assistant
supreethreddy.mamilla@wne.edu

WESTERN NEW ENGLAND UNIVERSITY WNE