

Incorporating Item Analysis Metrics into Programmatic Outcomes Assessment



Benjamin Shultz, Ph.D. and Rosalyn P. Vellurattil, Pharm.D., CHCP
University of Illinois Chicago College of Pharmacy (UIC COP)



KEY TAKEAWAYS

- The Office of Assessment can assist faculty improve item writing through item analysis metrics and competency levels that more accurately measure student achievement of desired knowledge.

OBJECTIVE

- The UIC COP uses ExamSoft® to administer exams and quizzes electronically throughout the PharmD curriculum.
- Existing literature provides guidance on when instructors should consider revising or removing items based on item statistics and review.¹⁻⁴ Limited information is available on performing score adjustments.
- The objective of this study is to improve programmatic outcomes at UIC COP by exploring item analysis metrics for incorporation into ExamSoft® tagging data.

METHODS

- Exam items from 21 core courses were tagged to outcomes in ExamSoft® and sorted into four competency bands based on percentage of correct answers.
 - <70% Below expectations
 - 70-79% Minimum expectations
 - 80-89% Meets expectations
 - >90% Exceeds expectations
- Items were filtered by difficulty (0.45-0.75) and point biserial (>0.20) criteria and sorted into competency bands based on standard deviations from the mean score.
- Outcomes with less than five tagged items were excluded.
- Revised competency bands were compared with previously tagged items of all difficulty levels and point biserial ranges.

RESULTS

Figure 1: Comparison of Performance

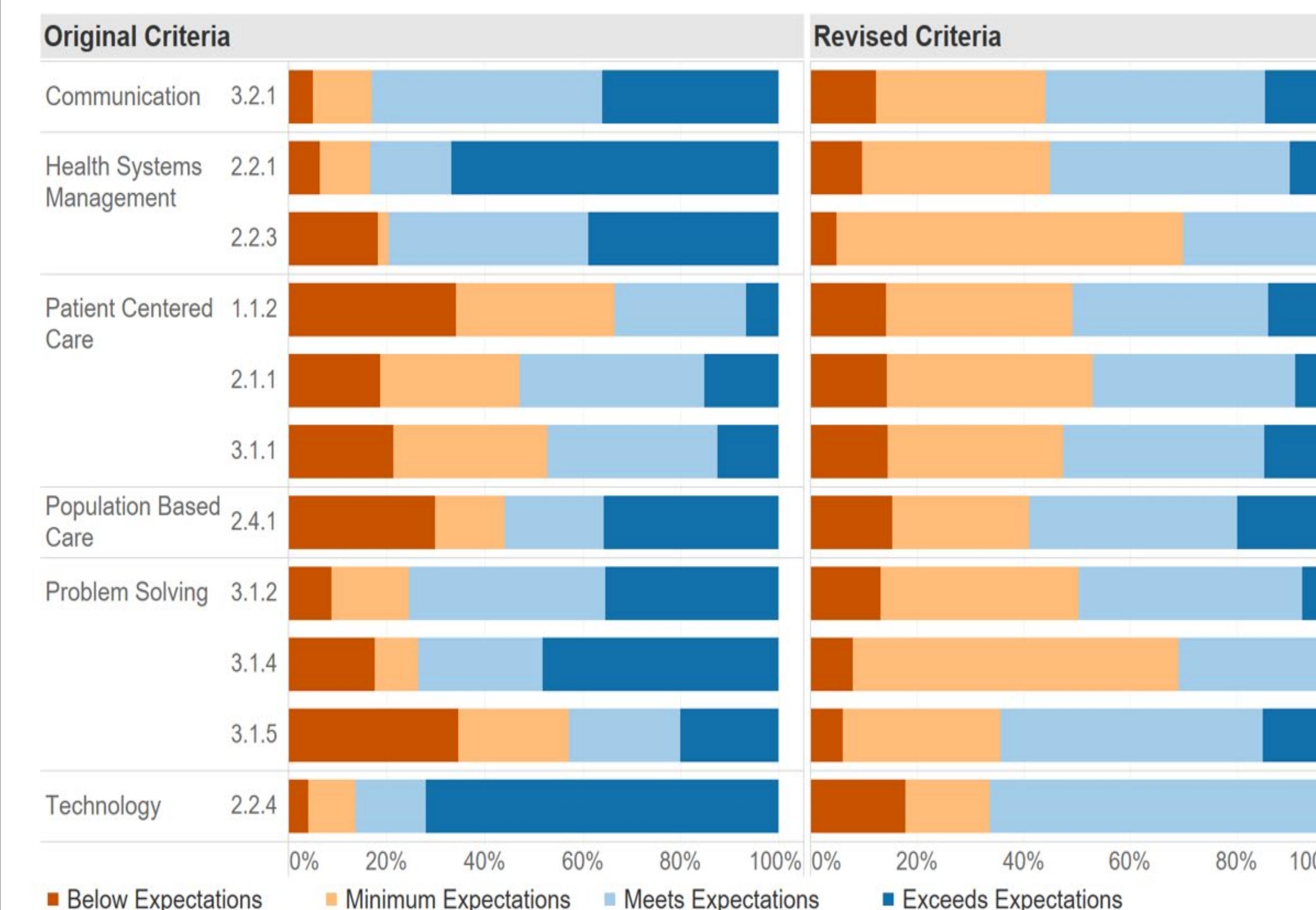


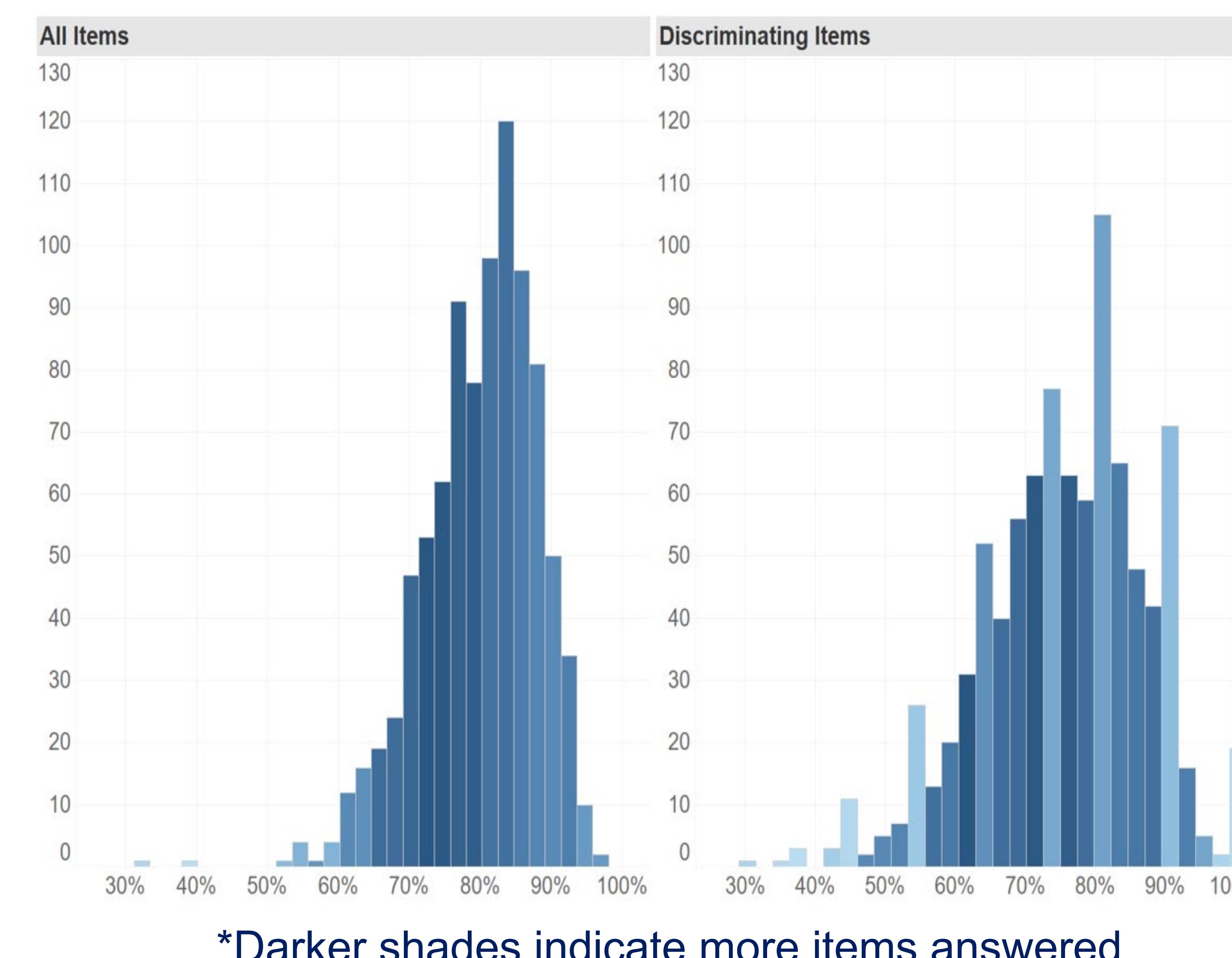
Table 1: Tagging Frequency

Questions Tagged Per Student by Outcome	Original Criteria	Revised Criteria
Communication	3.2.1	44 14
Health Systems Management	2.2.1	42 22
	2.2.3	39 8
Patient Centered Care	1.1.2	309 131
	2.1.1	226 96
	3.1.1	193 82
Population Based Care	2.4.1	12 11
Problem Solving	3.1.2	70 27
	3.1.4	38 6
	3.1.5	14 9
Technology	2.2.4	18 3
Total Questions Tagged		892 344

Table 2: Descriptive Statistics (N=907)

Statistic	All	Discriminating
Min. Score	28.60	27.30
Max Score	100.00	100.00
Mean Score	80.00	75.60
Std. Dev.	0.08	0.12
Skewness	-1.06	-0.54
Kurtosis	3.13	0.43

Figure 2: Distribution of Average Scores*



FINDINGS

- The original database included 892 questions tagged per student but decreased to 344 after applying exclusion criteria.
- Using standardized scores to set achievement criteria shows how students perform relative to each other and accounts for the greater difficulty of the items tagged.
- In both cases, students on the scoring extremes answered fewer questions.
- The revised criteria yields a wider, flatter distribution of scores by focusing on items that discriminate between the highest and lowest performers.
- The Office of Assessment will use these data to assist faculty with targeting exam items to meet desired item analysis criteria.
- Remediation efforts will specifically target students performing below expectations using this approach.
- Supportive efforts to assist students can also be identified and provided earlier on to help aid in student success.

REFERENCES

- Rudolph MJ, Daugherty KK, Ray ME, et al. Best practices related to examination item construction and post-hoc review. *Am J Pharm Educ.* 2018;83(7):7204.
- Ray ME, Daugherty KK, Lebovitz L, et al. Best practices on examination construction, administration, and feedback. *Am J Pharm Educ.* 2018;82(10):7066.
- Haladyna TM. *Developing and Validating Multiple-Choice Test Items.* 3rd ed. Routledge; 2004.
- Yudkowsky R, Park YS, Downing SM. *Assessment in Health Professions Education.* 2nd ed. Routledge; 2020.