The Reliability of Unproctored Online Precalculus Placement Test for Calculus Courses

Meiqin Li, University of Virginia

Dr. Meiqin Li is an Associate Professor at the University of Virginia. She joined UVA in 2017 after earning her Ph.D. in Applied Mathematics from Texas A&M University–College Station. Dr. Li has a strong interest in STEM education, particularly in curriculum development and redesign, the integration of instructional technologies to support student success, the creation of inclusive and diverse learning environments, and the use of a variety of pedagogical approaches to foster student learning. In addition to her education-focused work, her research interests include numerical computation, optimization, nonlinear analysis, and data science.

Dr. Hui Ma, University of Virginia

Dr. Hui Ma is an associate professor of applied math at the University of Virginia. She holds a Ph.D. in applied mathematics from the University of Alabama at Birmingham. Her current research interest focuses on mathematics education and STEM education, in particular student-focused instruction and learning-based grading.

Prof. Lindsay Wheeler

Dr. Lindsay Wheeler is Assistant Director of STEM Education Initiatives and Assistant Professor in the Center for Teaching Excellence at the University of Virginia. Dr. Wheeler develops and implements educational development programs and courses to suppor

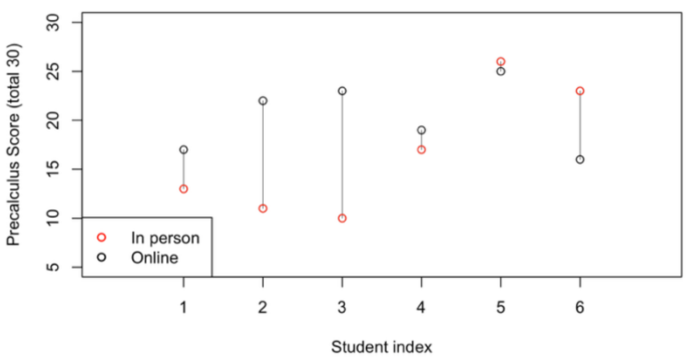
The Reliability of Online Un-proctored Precalculus Placement Test for Calculus Courses

1. Introduction

Appropriate placement into college-level mathematics courses is essential for student success and retention. Misplacement can lead to academic struggles, reduced confidence, and increased attrition [1], [2]. Since the pandemic, online assessments become more common for many subjects in higher education, and online placement tests for math courses such as ALEKS and Accuplacer are used more often as well. At our institution, to better meet our students' need and background, we designed our own placement test for math courses within the School of Engineering & Applied Science (SEAS). Incoming engineering students take an Online Un-proctored Placement Test (OUPPT) before enrolling in calculus courses. However, recent observations suggest that OUPPT scores may not reliably reflect students' actual preparedness. Here we study the reliability as measuring concepts consistently.

1.1. Motivation

In Summer 2024, six students were placed into Calculus II based on high OUPPT scores. Three of these students (n=1, 2,3) scored significantly lower on a comparable In-person Proctored Precalculus Assessment (IPPA) and struggled throughout the course. They should be placed into Calculus I based on IPPA. In contrast, the other three students, who performed similarly on the IPPA as in OUPPT (n=4, 5, 6), succeeded in Calculus II. This observation, along with past instances of struggling students in calculus courses motivates a critical question at our institution: How reliable are OUPPTs in appropriately assigning students into different courses?



This experience during Summer 2024 motivated us to extend the analysis to a larger scale by directly comparing results from the online un-proctored placement test (OUPPT) and the in-person proctored precalculus assessment (IPPA) to evaluate the reliability of OUPPT. The IPPA has previously been examined in [3] and was shown to be a strong indicator of whether students are likely to perform successfully in Calculus I or Calculus II. By using IPPA as a benchmark, this study aims to determine whether OUPPT yields comparable and consistent placement outcomes.

1.2. OUPPT and IPPA

Online Un-proctored Precalculus Placement Test (OUPPT) and In-person Proctored Precalculus Assessment (IPPA) are different but comparable versions of the same precalculus evaluation. Our research team developed a comprehensive list of precalculus learning objectives such as simplifying expressions, applying Laws of Logs, determining limits, etc. The OUPPT is regulated by such list of learning targets with 30 multiple choice problems, one point awarded for each multiple-choice problem. Students took OUPPT over the summer before their first semester at the engineering school. IPPA is meticulously designed to align with the specified learning objectives outlined in the OUPPT content and format (Table 1)[4], and students took IPPA at the beginning of the semester in class with instructors/TAs proctoring.

OUPPT	IPPA
Q10: Solve for x: $x^2 - 2x > 3$ A. $(-\infty, -1) \cup (3, \infty)$ B. $(-\infty, -2) \cup (-1, \infty)$ C. $(-1,3)$ D. $(-2, -1)$	Q10: Solve for x: $x^2 - 9x < -14$ A. $(-\infty, 9) \cup (14, \infty)$ B. $(-\infty, 2) \cup (7, \infty)$ C. $(2, \infty)$ D. $(2,7)$

Table 1: Alignments for the learning objective, "can solve quadratic inequalities (N)"

The mapping of each multiple-choice problem with the learning targets are displayed in Table 2.

PreCalculus Test Problem	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Learning Targets	S, I, E, L	S, L	S, E, L	S, P	S, E	QA	Q	Q	NA	N	T	T	T	T	T	TU	T, TD	T	T, TD	TD	LF	LF	LC	LC	LF	TG	TG	GP	GE	GL

Table 2: mapping of each multiple-choice problem with the learning targets, details refer to [5]

1.3. Literature Review

A substantial body of research has examined the predictive validity of calculus placement exams. Scott-Clayton's foundational study demonstrated that math placement exams are moderately effective in predicting success in college-level mathematics, especially when compared to placement in language courses, but they also produce nontrivial rates of misplacement [6]. Subsequent studies confirmed that placement scores correlate positively with student performance in Calculus I, particularly for engineering students, for whom early calculus success is tightly linked to persistence in the major [7]. Recent quantitative analyses using regression discontinuity designs have shown that students just below placement cutoffs often experience adverse outcomes when placed into prerequisites, suggesting that rigid thresholds may unnecessarily delay academic progress [8]. D. Bullock *et al.* investigated placement test design, predictive validity, and long-term impacts on retention and graduation [9].

In engineering education contexts, calculus placement is closely associated with student retention and persistence. Studies conducted at military, public, and research-intensive institutions indicate that failure to place directly into Calculus I increases the likelihood of major switching or attrition among engineering students [10], [11]. Ryan *et al.* found that math placement exams influence not only calculus enrollment but also longer-term engineering persistence, with differential impacts across gender and underrepresented minority groups [8]. These findings reinforce concerns that placement exams may unintentionally exacerbate inequities if cutoff scores are not carefully validated.

Research has also explored alternative and complementary indicators of calculus readiness. Standardized test scores such as SAT Math have demonstrated strong predictive power for success in first-year college mathematics and STEM courses, even after controlling for high school GPA [12]. Likewise, AP Calculus performance has been widely used as a placement proxy, though institutions increasingly recognize variability in instructional rigor across high schools [13] . As a result, many universities adopt multiple-measures placement models**,** combining placement tests, standardized exams, coursework history, and diagnostic assessments to improve accuracy and reduce severe misplacement [6], [12].

Another key theme in the literature is the diagnostic versus gatekeeping role of placement tests. Adaptive systems such as ALEKS emphasize identifying specific skill gaps and providing targeted learning modules prior to course enrollment [14]. Empirical evidence suggests that students who engage in structured remediation within these systems can significantly improve placement outcomes and subsequent calculus performance [8], [14]. This diagnostic approach aligns with calls from mathematics and engineering education researchers to shift placement away from exclusionary practices and toward academic support models that promote student success.

Recent research has increasingly focused on online and computer-adaptive mathematics placement tests as scalable tools for assessing calculus readiness among engineering students. Ferrer *et al.* examined the implementation of an adaptive online placement system within a large engineering college and reported that adaptive testing provided more precise diagnosis of deficiencies in algebra and trigonometry than traditional fixed-form exams, leading to improved initial course placement decisions [15]. Their findings suggest that online placement environments can better align engineering students with appropriate calculus pathways by adjusting item difficulty in real time.

The use of learning analytics embedded in online placement systems has also gained attention. Bowers and Zhou analyzed interaction data generated during online placement exams for first-year engineering students and found that behavioral indicators—such as time spent per item and revision frequency—were significant predictors of Calculus I performance beyond the final placement score [16]. This work highlights an advantage unique to online placement tests: the availability of rich process-level data that can inform both placement and early academic interventions.

Equity implications remain a central concern in the online testing literature. Relles and Tierney investigated the effects of online, un-proctored math placement exams on STEM-intended students and observed increased score dispersion across socioeconomic groups compared with in-person testing [17]. However, they also found that policies allowing multiple placement attempts combined with mandatory online remediation substantially reduced inequities in calculus placement outcomes. These results underscore that disparities in online placement are strongly influenced by institutional policy design rather than test modality alone.

The rapid expansion of remote online placement testing following the COVID-19 pandemic has prompted validation studies. Miller and Best compared pre-pandemic in-person placement exams with remotely administered online tests for engineering cohorts and found no significant reduction in predictive validity for calculus performance, provided assessments emphasized conceptual prerequisite knowledge rather than routine computation [18]. This evidence supports the continued use of online placement exams in post-pandemic engineering education contexts.

While placement test has been widely studied, among some focus on online placement tests, very limited attention has been given to the **reliability-measure concepts consistently-** of online un-proctored placement tests, especially in STEM disciplines. Existing studies that do address online placement reliability tend to focus on liberal arts contexts, such as language learning [19], leaving a gap in the literature regarding quantitative subjects.

1.4. Purpose & Research Questions

Our study aims to address this gap by systematically comparing overall and topic-specific scores obtained from highly similar online and in-person precalculus exams. The novelty of this work is several folds. First, we investigate the precalculus performance only, which serves as an important indicator of success in calculus course. Second, to the best of our knowledge, the reliability of online un-proctored precalculus placement tests for calculus courses has not yet been examined for engineering students**,** despite the widespread adoption of such assessments since the COVID-19 pandemic. Third, we compare highly similar assessments—an online un-proctored precalculus placement test (OUPPT) and an in-person proctored precalculus assessment (IPPA)—administered to the same students, allowing for a direct investigation of the reliability of OUPPT, an approach that has not been reported in the existing literature. Fourth, rather than focusing solely on performance-based outcomes, we broaden the scope of analysis by examining students' perceptions of the reliability of OUPPT and their views on the fairness and appropriateness of their placement under the current criteria**.**

This study contributes meaningfully to the literature on mathematics placement assessment and has practical implications for institutions that are considering or have already implemented online un-proctored placement tests.

Specifically, our research questions are:

RQ1. To what extent does the online un-proctored precalculus placement test (OUPPT) differ with in person proctored precalculus assessment (IPPA) in term of precalculus readiness for Calculus I and Calculus II?

RQ2. How consistent is the topic-level reliability of the OUPPT in assessing precalculus knowledge for Calculus I and II compared with IPPA? Are certain mathematical domains (e.g., functions, trigonometry, algebraic manipulation) more prone to score discrepancies between the online and in-person assessments?

RQ3. How do students perceive the OUPPT and IPPA and their placements under the current criteria? What, if any, differences exist in OUPPT, IPPA, and final exam scores for students based upon their perceptions of their placement?

1.5. Context

This IRB approved study occurred at a large, research-intensive, public, predominately white institution (PWI) in the Mid-Atlantic region of the U.S and was supported by HHMI IE3 community-The University of Virginia. At our institution, students apply to and are accepted directly into our Engineering School. All engineering students are required to have credits for APMA 1110-Single Variable Calculus II and APMA 2120-Multivariable Calculus as part of the engineering curriculum, regardless of major. Starting from Fall 2025, all engineering students are required to have credits for APMA 1090-Single Variable

Calculus I as well. The Center for Applied Mathematics (APMA) within the Engineering School teaches a variety of math courses to both graduate and undergraduate students from Engineering. Almost all students in these Math courses are pursuing engineering degrees, with a few students enrolled from outside Engineering.

Because of the math requirements for engineering majors, APMA administers OUPPT to all incoming engineering students prior to the beginning of the Fall semester each year. Along with students' AP calculus scores obtained from high school, OUPPT helps to place students at the best calculus level for them.

All students are required to have credit for APMA 1110-Calculus II and APMA 2120-Calculus III. Before Fall 2025, students were not required to have APMA 1090-Calculus I credit and the OUPPT played an important role in placing students into either Calculus I or Calculus II. Students who achieved only the threshold OUPPT scores were placed into Calculus II, while those who scored below the threshold were placed into Calculus I. Beginning in Fall 2025, all incoming engineering students are required to have credit for Calculus I in addition to Calculus II & III. As a result, the impact of the OUPPT on placement decisions has been reduced, although it still plays an important role in making placement recommendations.

Given the importance of precalculus skills for success in calculus, a mastery-based precalculus component was incorporated into both Calculus I and Calculus II in Fall 2024 and Fall 2025 to reinforce students' foundational knowledge. To establish a baseline measure of students' precalculus proficiency, all students completed the in-person proctored precalculus assessment (IPPA) at the very beginning of the semester. The IPPA was administered prior to instruction and under proctored conditions. [3] found that IPPA is a reliable indicator of students' incoming precalculus preparedness and is a strong predictor of students' calculus performance.

2. Study Design and Methods

2.1. Participants

The study included students enrolled in APMA 1090 (Calculus I) and APMA 1110 (Calculus II) during Fall 2024 and Fall 2025. In Fall 2024**,** APMA 1090 enrolled approximately 170 students across five sections taught by three instructors, while APMA 1110 enrolled approximately 280 students across seven sections taught by four instructors. In Fall 2025**,** APMA 1090 included approximately 230 students across seven sections taught by five instructors, and APMA 1110 included approximately 180 students across four sections taught by three instructors**.** In total, the dataset comprises approximately 860 students distributed across 23 course sections and taught by many different instructors.

2.2. Data Sources & Collection

The primary data sources include Fall 2024 & Fall 2025 APMA 1090-Calculus I students' OUPPT scores, IPPA scores, and final exam scores. Fall 2024 & Fall 2025 APMA 1110-Calculus II students' OUPPT scores, IPPA scores, and final exam scores. Fall 2025 end-of-semester survey for APMA 1090 and APMA 1110. Four students' interview data. The end-of-semester survey was administered to students at the end of Fall 2025 via Qualtrics. It comprised Likert scale open-ended questions aimed at gauging students' perceptions of the OUPPT and IPPA and their placements under the current criteria.

2.3. Study Methods

For quantitative data, we utilize a mix of statistical tools such visualizations, descriptive tables, paired t-test, ANOVA tests, and non-parametric tests to analyze the data to draw conclusion. For qualitative data, we use both deductive and inductive approaches to code the responses and draw conclusion. For RQ1, we analyze overall score differences using thresholds of ±10%, ±15%, ±20%, and ±25% and investigate the Pearson correlation coefficients of OUPPT and IPPA. For RQ2, we examine individual test problem s by topic to assess discrepancies. For RQ3, we use mixed method to analyze survey and interview responses by various statistical tools. For example, we conducted an ANOVA to identify differences in OUPPT, IPPA, and final exam scores between students who held different perceptions of their placement accuracy (i.e., too easy, accurate, too hard). Due to the small sample size of students who identified Calc I or Calc II as too hard, these students were excluded from the analysis. We also conducted Pearson correlational analyses to identify relationships between OUPPT, IPPA, and final exam scores for students based on their perceptions of placement accuracy.

3. Data Analysis and Results

This section examines the data and findings from four cohorts: APMA 1090 (Calculus I) in Fall 2024, APMA 1110 (Calculus II) in Fall 2024, APMA 1090 in Fall 2025, and APMA 1110 in Fall 2025. In all figures and tables, "PT" indicates the Online Un-proctored Precalculus Placement Test (OUPPT), while "1st PA" denotes the In-person Proctored Precalculus Assessment (IPPA). To accommodate space limitations, all visualizations are presented in a condensed format; readers may need to enlarge them for better readability.

3.1. RQ1: To what extent does the online un-proctored placement test (OUPPT) differ with IPPA in term of precalculus readiness for Calculus I and Calculus II?

Overall Discrepancy. Figure 1 displays overall discrepancies between OUPPT and IPPA scores, with the x-axis representing student indices ordered from lowest to highest IPPA score (maximum score = 30). Across all panels, the plots show a noticeable vertical spread of OUPPT scores around the ordered IPPA scores, indicating meaningful individual-level discrepancies between the two assessments. These discrepancies are most pronounced among students in the low-to-upper middle range of the score distribution rather than at the highest ends.

Differences by course level are also evident. P In Calculus II, a substantially larger proportion of students exhibit higher OUPPT scores than IPPA scores, with relatively few students showing the opposite pattern. In contrast, while many Calculus I students also have higher OUPPT scores than IPPA, there remains a notable number of students whose OUPPT scores are lower than their IPPA scores. This suggests greater discrepancy and asymmetry between the two placement measures for Calculus II students compared with those enrolled in Calculus I.

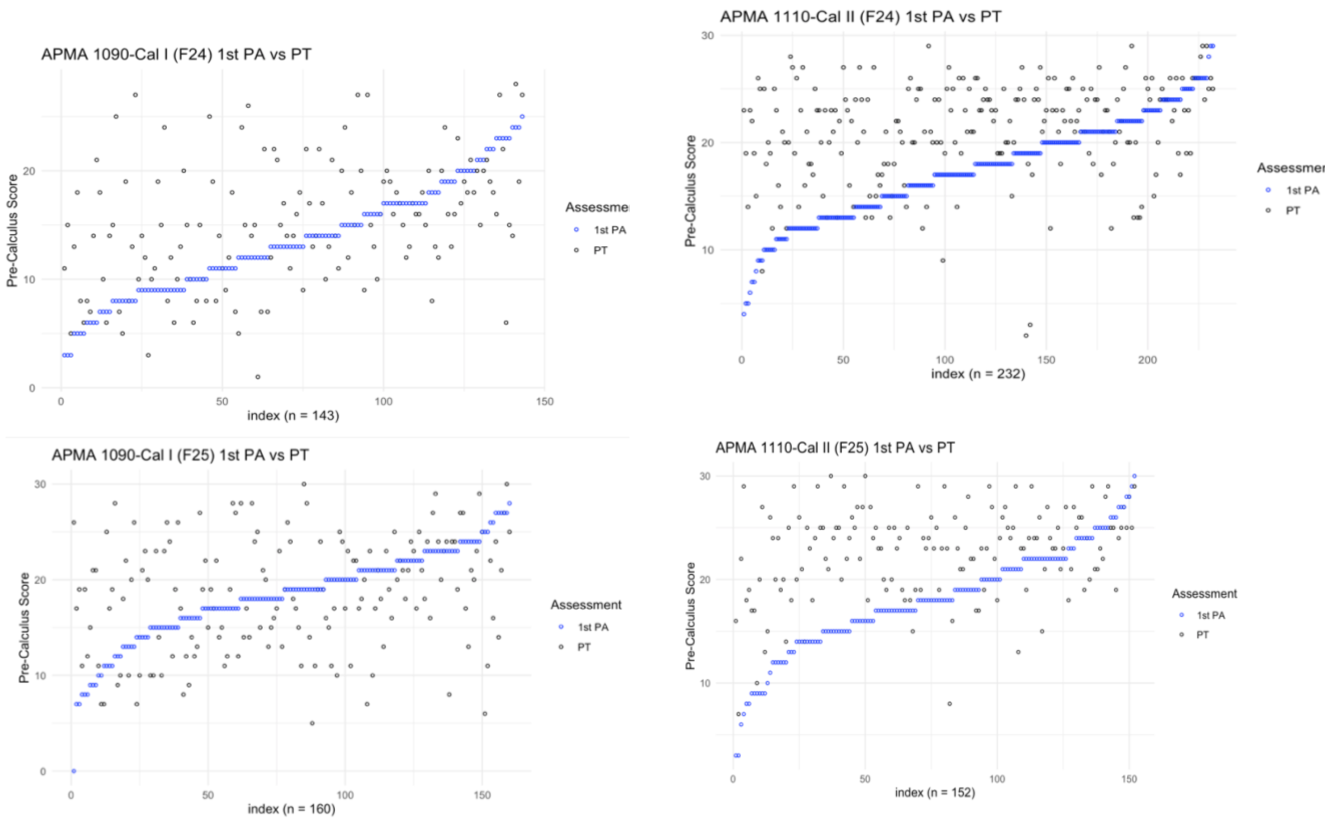


Figure 1: Overall discrepancy between OUPPT and IPPA

Paired t-Test. We further applied paired t-tests across all four cohorts to compare OUPPT and IPPA scores. The results showed that the differences between OUPPT and IPPA were statistically significant ($p \approx 0$) for both Fall 2024 and Fall 2025 Calculus II cohorts, with OUPPT scores outperforming IPPA. These findings are consistent with and reinforce the patterns observed in the visualizations.

For this analysis, we initially treated each cohort as a single aggregated dataset, without disaggregating by IPPA performance levels (e.g., IPPA<40%, 40%<IPPA<60%, etc.). Under this aggregated approach, no statistically significant difference ($p > 0.1$) was observed between OUPPT and IPPA for the Fall 2024 and Fall 2025 Calculus I cohorts. However, for such two cohorts, when the data were disaggregated based on IPPA performance categories (late IPPA <40%, 40% ≤ IPPA < 60%, 60% ≤ IPPA < 70%, 70% ≤ IPPA < 80%, and IPPA ≥ 80%), statistically significant differences ($p<0.01$) between OUPPT and IPPA emerged for the lower-performing groups, specifically those with IPPA <40% and 40% ≤ IPPA < 60%.

Discrepancy at ±20%, ±25% and ±30% . Figure 2 presents discrepancies between OUPPT and IPPA scores under increasing tolerance thresholds (±20% to ±30%), treating IPPA as a benchmark and framing the comparison within a test–retest context in which some score variation is expected. Across all four cohorts, many students' OUPPT scores fall within narrower variation bands; however, even after allowing for a relatively large ±30% variation, a nontrivial number of students remain outside these bounds.

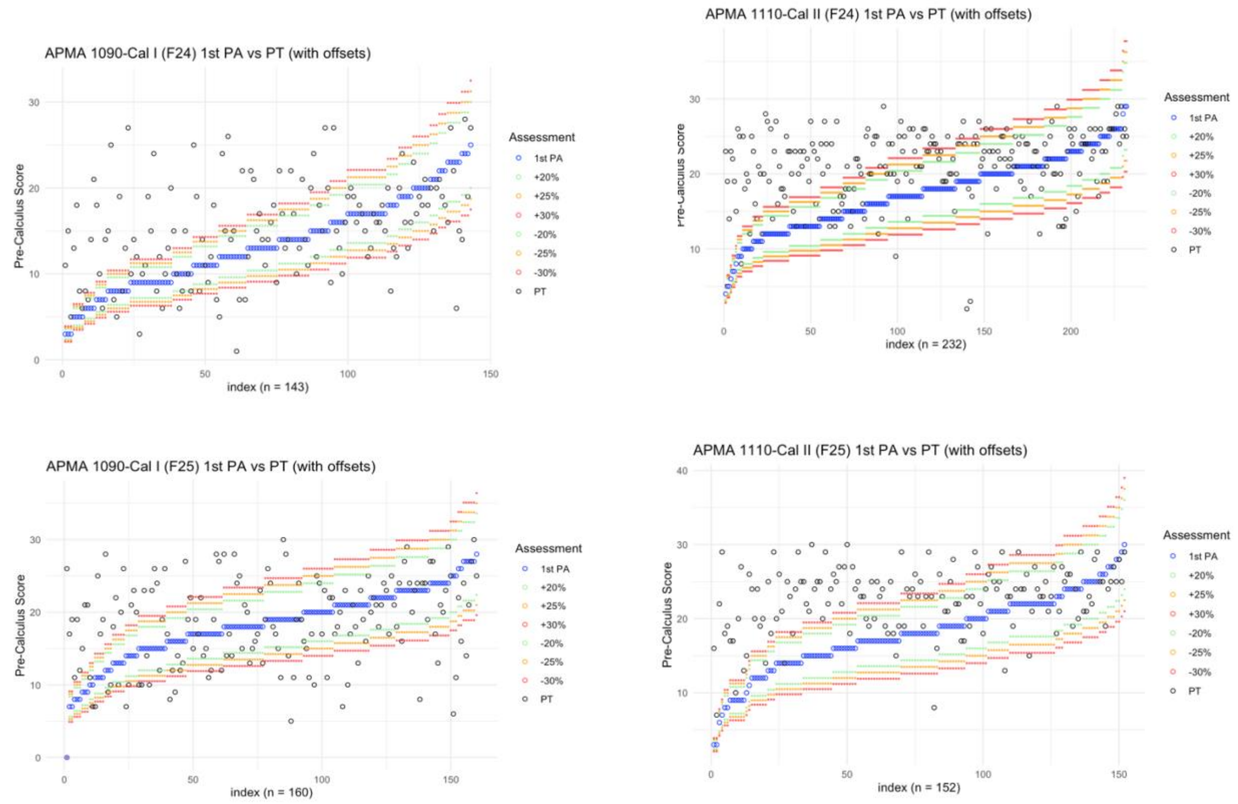


Figure 2: Discrepancy between OUPPT and IPPA at ±20% to ±30% variations

This pattern is most evident among Calculus II students, for whom OUPPT scores are more frequently higher than corresponding IPPA scores**,** with relatively few instances of lower OUPPT scores. In contrast, Calculus I cohorts continue to show discrepancies in both directions, with a meaningful number of students scoring lower on OUPPT than on IPPA. While these results do not imply systematic inflation by the online assessment, they indicate that discrepancies exceeding typical test–retest variability persist, particularly for Calculus II placements. As a result, reliance on OUPPT alone may carry greater placement decision risk for students near Calculus II cutoffs compared with those near Calculus I thresholds.

Based on the observed score discrepancies between OUPPT and IPPA, students were further grouped according to their IPPA (1st PA) performance levels to identify which score ranges exhibited the greatest variation. Specifically, students were categorized into five groups: IPPA < 40%, 40% ≤ IPPA < 60%, 60% ≤ IPPA < 70%, 70% ≤ IPPA < 80%, and IPPA ≥ 80%. A ±25% difference was selected as the maximum acceptable variation under a test–retest framework, and descriptive statistics summarizing agreement within this threshold are reported in Table 1**.**

Table 3 presents the number and proportion of students whose OUPPT scores fall within ±25% of their corresponding IPPA scores, stratified by IPPA score group and cohort. Across all four cohorts, approximately 50–54% of students had OUPPT scores within the ±25% threshold, indicating moderate overall agreement between the two assessments. However, agreement varied substantially by IPPA score range. In both Calculus I and Calculus II cohorts, students with higher IPPA scores (≥70%) consistently demonstrated higher within-25% agreement, often exceeding 80–95%, whereas students in the lower and middle IPPA ranges (<60%) showed markedly lower agreement rates**.** This pattern is especially pronounced in Calculus II cohorts, where fewer than 40% of students scoring below 60% on IPPA had OUPPT scores within ±25%.

APMA 1090-Cal I (F24) Count PT Scores Within 25% 1st PA

Group	Within 25%	PT within 25% 1st PA	PT NOT within 25% 1st PA
1st PA < 40%	33.3%	17	34
40% ≤ 1st PA < 60%	58.6%	34	24
60% ≤ 1st PA < 70%	66.7%	10	5
70% ≤ 1st PA < 80%	63.6%	7	4
1st PA ≥ 80%	75.0%	3	1
TOTAL	51.1%	71	68

APMA 1110-Cal II (F24) Count PT Scores Within 25% 1st PA

Group	Within 25%	PT within 25% 1st PA	PT NOT within 25% 1st PA
1st PA < 40%	15.8%	3	16
40% ≤ 1st PA < 60%	37.5%	33	55
60% ≤ 1st PA < 70%	50.0%	25	25
70% ≤ 1st PA < 80%	78.9%	30	8
1st PA ≥ 80%	96.0%	24	1
TOTAL	52.3%	115	105

APMA 1090-Cal I (F25) Count PT Scores Within 25% 1st PA

Group	Within 25%	PT within 25% 1st PA	PT NOT within 25% 1st PA
1st PA < 40%	7.1%	1	13
40% ≤ 1st PA < 60%	35.6%	16	29
60% ≤ 1st PA < 70%	55.8%	24	19
70% ≤ 1st PA < 80%	83.8%	31	6
1st PA ≥ 80%	68.4%	13	6
TOTAL	53.8%	85	73

APMA 1110-Cal II (F25) Count PT Scores Within 25% 1st PA

Group	Within 25%	PT within 25% 1st PA	PT NOT within 25% 1st PA
1st PA < 40%	11.1%	1	8
40% ≤ 1st PA < 60%	20.8%	11	42
60% ≤ 1st PA < 70%	51.6%	16	15
70% ≤ 1st PA < 80%	81.5%	22	5
1st PA ≥ 80%	95.7%	22	1
TOTAL	50.3%	72	71

Table 3: Count of OUPPT Scores Within ±25% of IPPA Scores Across All Cohorts

To complement the numerical summaries in Table 3, Figure 3 presents bar plots corresponding to the four cohorts, displaying the proportion of students whose OUPPT scores fall within ±25% of their IPPA scores across IPPA performance groups. These bar plots provide a visual comparison of agreement patterns across score ranges, course levels, and cohorts that is less apparent from tabular counts alone.

Consistent with the tabulated results, the bar plots show a systematic increase in within-25% agreement as IPPA score level increases. In all cohorts, agreement rates are lowest among students with IPPA scores below 60% and rise substantially for students scoring 70% or higher**.** This upward trend is particularly clear in the Calculus II cohorts, where agreement among high-scoring students reaches nearly complete alignment, while agreement among low-scoring students remains relatively limited. Calculus I cohorts exhibit similar patterns, though with less pronounced contrasts between adjacent IPPA score groups.

Figure 3 reinforces the descriptive findings from Table 3 by illustrating that convergence between OUPPT and IPPA scores is strongly associated with IPPA performance level. The consistency of these trends across cohorts suggests that the observed variation reflects systematic differences across score ranges rather than cohort-specific effects, providing additional context for understanding assessment alignment across the precalculus readiness spectrum.

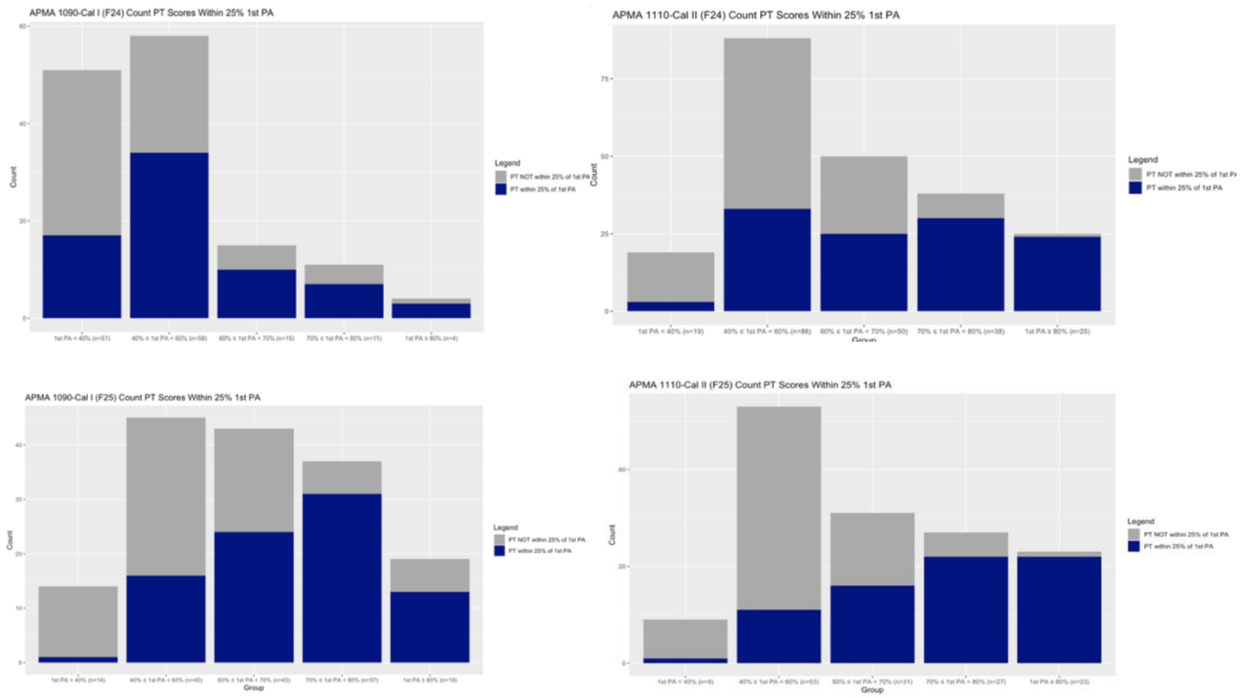


Figure 3: Bar plots of Count of OUPPT Scores Within ±25% of IPPA Scores Across All Cohorts

<u>*Correlation Analysis.*</u> Test-retest reliability is a psychometric measure that evaluates the temporal stability and consistency of an instrument by administering the same test to the same group of individuals on two or more separate occasions [20]. This metric, often called the coefficient of stability, is typically calculated using the Pearson Correlation Coefficient or the Intraclass Correlation Coefficient (ICC), where a value of 0.70 or higher generally indicates acceptable reliability [21]. The reliability of a measure is essential for ensuring that observed changes in scores are due to actual shifts in the underlying construct rather than random measurement error [22] [22].

	Fall24 APMA 1090-Cal I	Fall24 APMA 1110-Cal II	Fall25 APMA 1090-Cal I	Fall25 APMA 1110-Cal II
Pearson Correlation Coefficients between 1st PA and PT	0.4180064	0.2688672	0.2148673	0.3348183
Correlation test p value	p=3.053e-07	p = 5.348e-05	p= 0.006705	p= 4.364e-05

Table 4: Correlation Coefficients

The OUUPT and IPPA align naturally with a test–retest framework due to their design, including close problem-by-problem alignment targeting the same precalculus learning objectives. Accordingly, we conducted correlation analyses to examine the consistency between students' OUUPT and subsequent IPPA performance. The results are summarized in Table 4. Across both Calculus I and Calculus II and in both Fall 2024 and Fall 2025 cohorts, the observed Pearson correlation coefficients were statistically significant but fell well below the conventional threshold for acceptable test–retest reliability (>=0.7). These relatively weak correlations suggest limited temporal stability between the OUUPT and IPPA, providing further evidence that OUUPT scores may not reliably reflect students' true precalculus proficiency.

Summary. These results indicate that while OUPPT and IPPA produce more consistent signals for students with strong precalculus preparation, substantial discrepancies exceeding typical test–retest variation (±25%.) remain common for lower- and mid-performing students (~ half of all the students), particularly for those placed into Calculus II. This finding highlights increased placement uncertainty near key decision thresholds and motivates further examination of domain-level discrepancies in subsequent analyses.

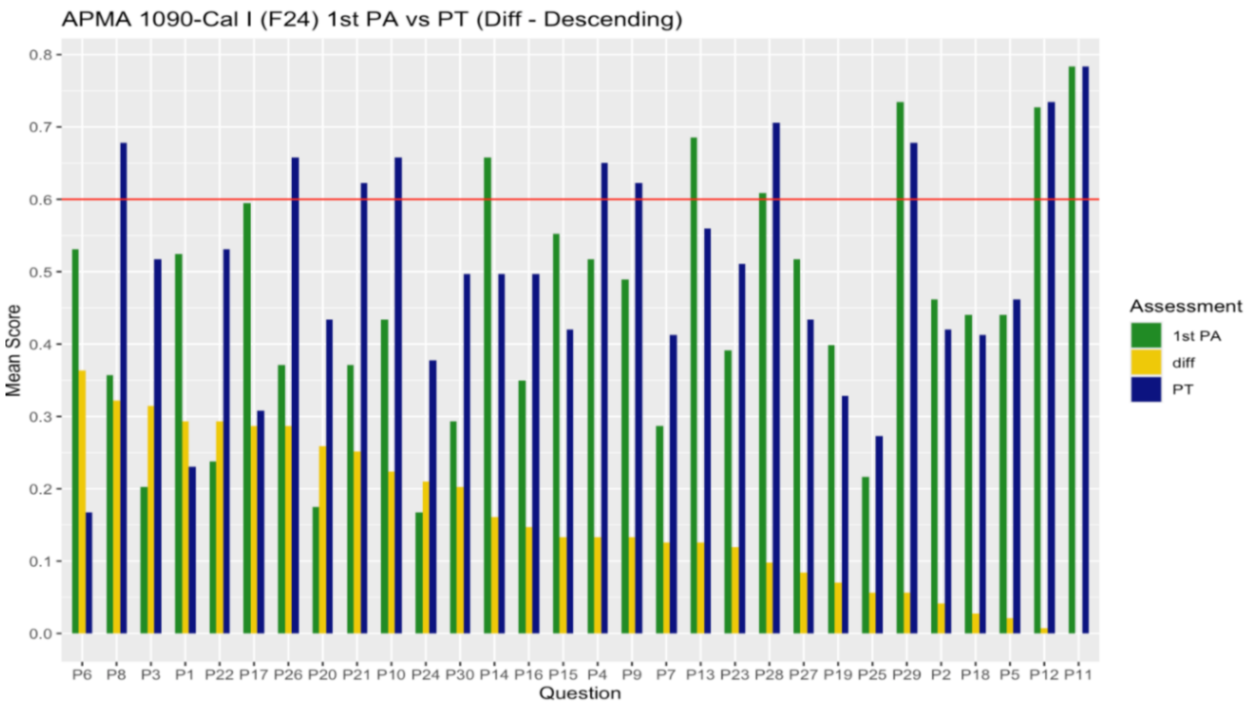
Several factors may contribute to the observed discrepancies between OUPPT and IPPA scores. The un-proctored nature of OUPPT may allow access to external resources or assistance, potentially leading to higher scores for some students, especially for students who may not have a strong background in precalculus, compared with the proctored IPPA. Differences in testing conditions—such as flexible timing, reduced test anxiety, or the ability to consult materials—may also affect performance. Additionally, greater variability among lower- and mid-performing students may reflect normal measurement error or less stable content mastery, while larger discrepancies in Calculus II cohorts could stem from higher placement stakes or uneven preparation across prerequisite topics.

3.2. RQ2: How consistent is the topic-level reliability of the OUPPT in assessing precalculus knowledge for Calculus I and II compared with IPPA? Are certain mathematical domains (e.g., functions, trigonometry, algebraic manipulation) more prone to score discrepancies between the online and in-person assessments?

In this section, we investigate the consistency of topic-level reliability of the online un-proctored precalculus placement test (OUPPT) in assessing precalculus knowledge for Calculus I and Calculus II, compared with the in-person proctored precalculus assessment (IPPA). The set of precalculus topics (learning targets) is provided in [5], and the alignment between OUPPT (or IPPA) test items and learning targets is summarized in Table 2. To examine topic-level discrepancies between the two assessments, we constructed bar plots for all four cohorts, ordering topics by the absolute score difference |OUPPT − IPPA| and, separately, by ascending IPPA performance. These visualizations were used to identify mathematical domains that exhibit larger or more systematic differences between the online and in-person testing formats.

Because the complete set of bar plots requires substantial space, we present as an illustrative example only the plots for the Fall 2024 APMA 1090 (Calculus I) cohort. Using this cohort, we demonstrate the analytic approach employed for all cohorts. The summarized descriptive interpretations for all cohorts are provided in this section, while the remaining cohorts' bar plots along with descriptive interpretation are included in Attachment 2.

Fall 2024 APMA 1090 (Calculus I) cohort. Figure 4 presents bar plots comparing mean problem-level performance on OUPPT and IPPA for the Fall 2024 APMA 1090 (Calculus I) cohort. Each problem is mapped to a specific precalculus learning target as summarized in Table 2, allowing discrepancies between the two assessments to be examined at the topic level. The upper panel orders items by descending absolute difference | OUPPT − IPPA |, highlighting learning targets with the largest modality-based discrepancies, while the lower panel orders items by ascending IPPA accuracy to identify topics associated with lower and higher levels of demonstrated proficiency under proctored conditions.



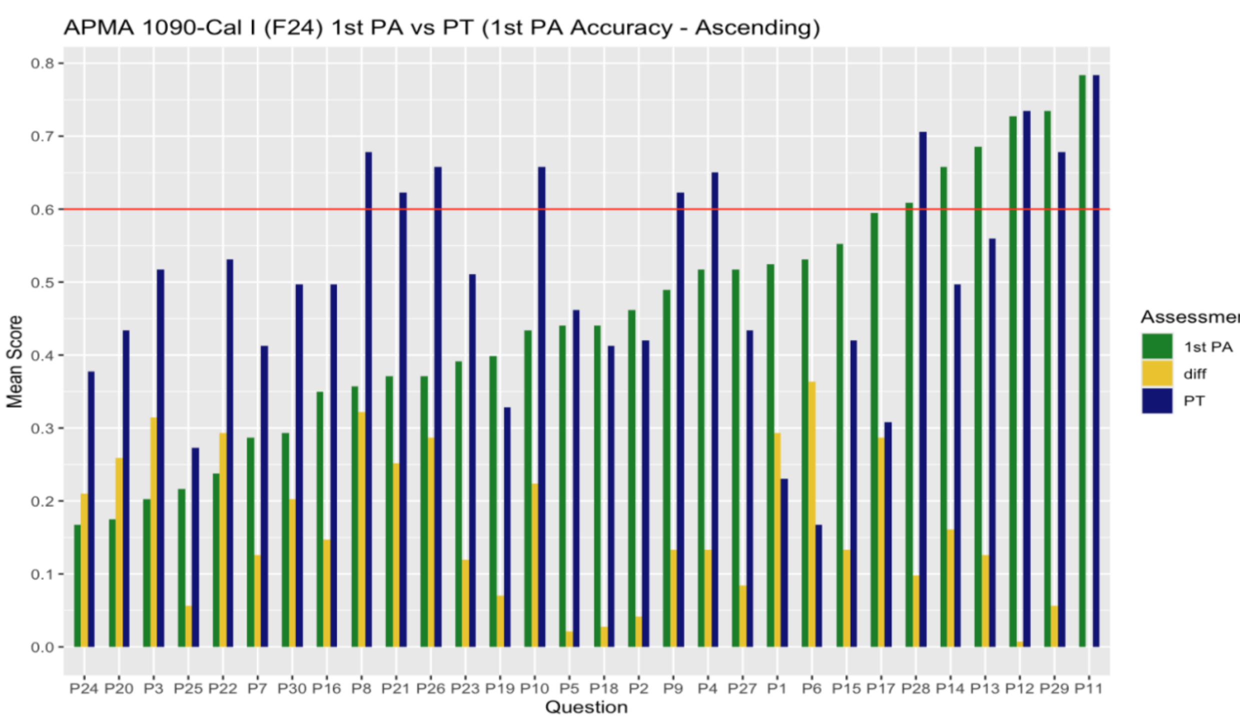


Figure 4: Bar plots of Individual Problems of IPPA Scores for Fall24 Cal I Cohorts

Across items with the largest discrepancies, differences between OUPPT and IPPA are concentrated primarily in algebra-focused learning targets, including simplifying expressions (S), laws of exponents and logarithms (E, L), polynomial division (P), solving equations and inequalities (Q, N), and absolute-value equations and inequalities (QA, NA). For many of these items, OUPPT mean scores exceed corresponding IPPA scores, indicating greater divergence between performance in online versus in-person testing environments for foundational algebraic skills. Items associated with limits and continuity (LF, LC) also appear among those with relatively large discrepancies and lower IPPA accuracy, suggesting both increased difficulty and greater variability across assessment modalities for these topics.

In contrast, items demonstrating the smallest discrepancies between OUPPT and IPPA are largely associated with trigonometry and function-graphing learning targets, including quick evaluation of trigonometric values (T), use of the unit circle (TU), trigonometric identities and graphs (TD, TG), and graphing power, exponential, and logarithmic functions (GP, GE, GL). These topics appear toward the right of the discrepancy-ordered plot and near the top of the accuracy-ordered plot, indicating comparatively strong performance and close alignment between the two assessments.

Overall, for the Fall 2024 Calculus I cohort, topic-level analysis shows that algebraic manipulation and early calculus concepts (limits and continuity) are more prone to score discrepancies between OUPPT and IPPA, whereas trigonometry and graphical interpretation of functions exhibit greater consistency across testing formats. These patterns provide evidence that agreement between the two assessments varies meaningfully by mathematical domain and motivate further examination of whether similar trends persist across cohorts and for Calculus II.

Paired t-Test. For each question, we further compared OUPPT and IPPA scores using paired t-tests. Based on the barplot ordered by descending differences between IPPA and OUPPT, it was not necessary to conduct all thirty paired t-tests individually. Instead, for example, in the Fall 2024 Calculus I cohort, problem P23 was the first to reach statistical significance ($p < 0.01$) (1st graph in Figure 4). Consequently, all problems appearing before P23 in the descending difference barplot can also be considered to exhibit statistically significant differences between OUPPT and IPPA. Although different cohorts showed slight shifts in the ordering of differences and in the specific threshold problem where statistical significance first appeared, all cohorts displayed consistent overall patterns as discussed above.

Summary. Across all four cohorts (Fall 2024 and Fall 2025; Calculus I and Calculus II), topic-level analyses reveal consistent domain-specific patterns in alignment between OUPPT and IPPA. Items associated with algebra-focused learning targets—including simplifying expressions (S), laws of exponents and logarithms (E, L), polynomial division (P), solving equations and inequalities (Q, N), and absolute-value equations and inequalities (QA, NA)—exhibit the largest and most frequent OUPPT–IPPA discrepancies. Learning targets related to limits and continuity (LF, LC) also tend to show lower IPPA accuracy and relatively larger differences between the two assessment modalities.

In contrast, items linked to trigonometric and function-graphing learning targets, such as quick evaluation of trigonometric values (T), use of the unit circle (TU), trigonometric graphs (TG), and graphing of power, exponential, and logarithmic functions (GP, GE, GL), consistently demonstrate higher accuracy and smaller score discrepancies across cohorts. While these domain-level trends are evident in both Calculus I and Calculus II, they are more pronounced in Calculus II**,** where OUPPT scores more frequently exceed IPPA scores across algebraic and limits-based topics. Overall, these findings indicate that topic-level consistency between online and in-person assessments varies systematically by mathematical domain and course level, providing contextual insight into observed placement score differences beyond overall test performance.

3.3. RQ3. How do students perceive the OUPPT and IPPA and their placements under the current criteria? What, if any, differences exist in OUPPT, IPPA, and final exam scores for students based upon their perceptions of their placement after completing the course?

The data used in this section consist of a survey to cohorts F25 APMA 1090-Cal I and F25 APMA1110-Cal II and interviews with four students (we did not administer the survey in Fall 24). The survey included two open-ended questions. This section is divided into three subsections. Section 3.3.1 examines students' perceptions of the OUPPT and IPPA, section 3.3.2 examines students' perceptions of their placement outcomes under the current placement criteria and section 3.3.2 explores whether differences exist in OUPPT, IPPA, and final exam scores according to students' perceptions of their placement after completing the course.

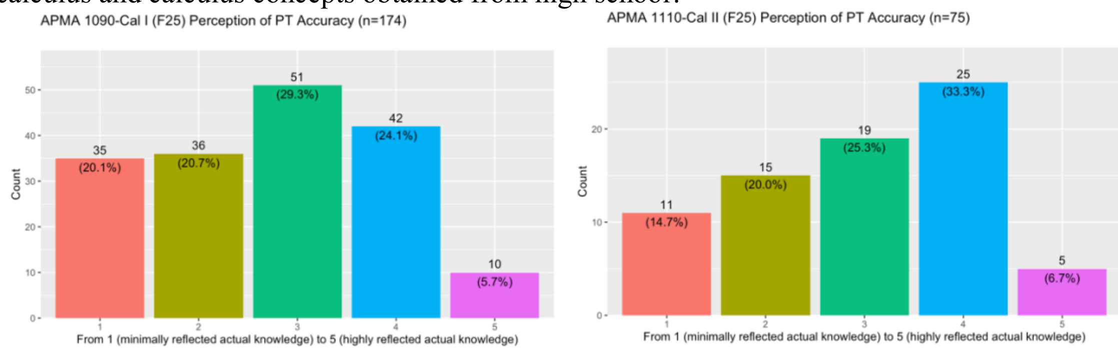
Before presenting the data analysis and results for RQ3, we note that these analyses reflect only students enrolled in Fall 2025, a term that implemented different placement criteria. Prior to Fall 2025, as described in Section 1.5, engineering students were not required to have credit for Calculus I, and the OUPPT played a central role in placing students into either Calculus I or Calculus II. Students who met only the threshold OUPPT score were placed into Calculus II, whereas those scoring below the threshold were placed into Calculus I. Findings from RQ1 and RQ2 strongly indicate that OUPPT was not a reliable placement tool, as many students—particularly those enrolled in Calculus II—earned substantially higher OUPPT scores than IPPA scores. Consequently, some students were placed into Calculus II based primarily on high OUPPT performance, regardless of their AP preparation, despite being insufficiently prepared for the course. This misalignment likely contributed to the high number of struggling students observed in Calculus II prior to Fall 2025.

Starting from Fall 2025, the criteria changed. In addition to the Calculus II and Calculus III credits, all engineering students are required to have Calculus I credit as well. The current criteria are: A score of 5 on the AP Calculus BC exam grants credit for both Calculus I and Calculus II, while a score of 4 on the BC exam grants credit for Calculus I only. Credit for Calculus I is also awarded to students who earn a score of 5 on the AP Calculus AB exam or a sub-score of 5 on the AB portion of the BC exam. Additionally, a score of 5 or higher on the IB Calculus examination grants credit for Calculus I. Students are placed according to the credit they receive from high school; however, placement notification emails include a note encouraging students to consider enrolling in a lower-level course if they do not pass OUPPT.

Accordingly, the RQ3 data analysis primarily reflects Fall 2025 students' placement perceptions under the current criteria, which substantially reduced the influence of the OUPPT. This analysis is nevertheless critical, as it provides the first opportunity (Fall 2025 was the first semester to implement the revised criteria) to evaluate whether the revised placement criteria appropriately align students with their course levels.

3.3.1. How do students perceive the OUPPT and IPPA and their placements under the current criteria?

Survey Q1: On a scale of 1–5, how accurately do you think the un-proctored precalculus placement test (OUPPT) you took during the summer before coming to UVA reflected your actual understanding of precalculus and calculus concepts obtained from high school?



APMA1090 mean=2.75 APMA 1110 mean=2.91

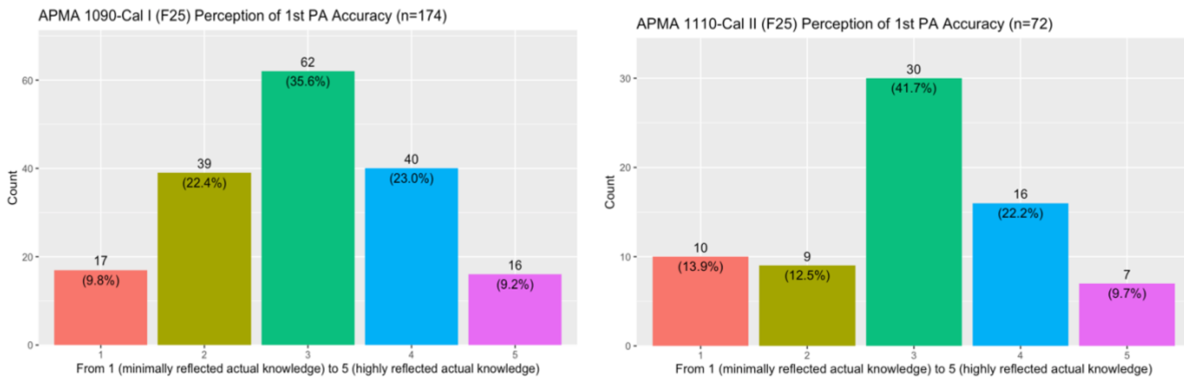
Figure 5: Students' Perception of OUPPT

Student responses indicate that the OUPPT was viewed as moderately accurate, with distributions centered near the scale midpoint in both courses. In APMA 1090 (Calculus I; $n = 174$), 40.8% rated accuracy as low (1–2), 29.3% as neutral (3), and 29.9% as high (4–5); the mean rating was 2.75 (95% CI: 2.57–2.92). In APMA 1110 (Calculus II; $n = 75$), perceptions were slightly more favorable: 34.7% rated low (1–2), 25.3% neutral (3), and 40.0% high (4–5); the mean was 2.97 (95% CI: 2.71–3.24). Only a small fraction in either course selected the maximum rating of 5 (5.7% in APMA 1090; 6.7% in APMA 1110). The between-course difference in means ($\Delta = 0.23$ on a 5-point scale) corresponds to a small effect (Cohen's $d \approx 0.19$), suggesting modestly higher confidence in the test's accuracy among Calculus II students. Overall, students tended not to view the un-proctored test as a highly accurate reflection of their prior precalculus/calculus knowledge, though many regarded it as at least somewhat informative (Figure 5).

Survey Q2. On a scale of 1–5, how accurately do you think the first in-person proctored precalculus assessment you took during your calculus class in F25 reflected your actual understanding of precalculus and calculus concepts obtained from high school?

Students perceived the IPPA as a moderately accurate reflection of their high-school precalculus and calculus preparation, with response distributions centered near the midpoint of the 5-point scale in both courses.

In APMA 1090 (Calculus I; $n = 174$), more than one-third of students selected the midpoint rating (3; 35.6%), while nearly equal proportions rated the assessment as low accuracy (1–2; 32.2%) and high accuracy (4–5; 32.2%). The mean perception score was 2.99, indicating that, on average, students viewed the IPPA as neither clearly inaccurate nor highly accurate. Notably, fewer than 10% of students selected the highest rating (5), suggesting limited strong endorsement of the assessment's accuracy.



APM1090 mean=2.99 APM1110 mean=3.01

Figure 6: Students' Perception of IPPA

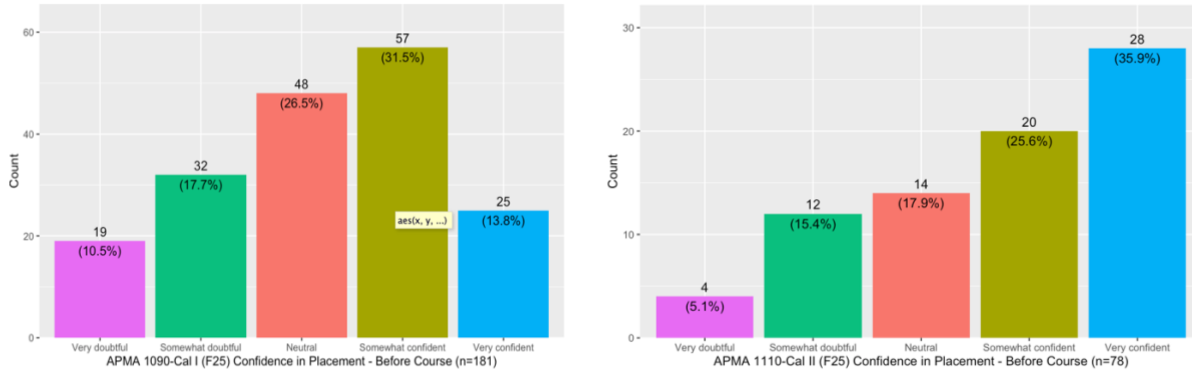
Similarly, in APMA 1110 (Calculus II; $n = 72$**)**, perceptions clustered even more strongly at the midpoint, with 41.7% of students rating the IPPA as a 3. Lower-accuracy ratings (1–2) accounted for 26.4% of responses, while 31.9% rated the assessment as highly accurate (4–5). The mean rating **(**3.01**)** was nearly identical to that of APMA 1090, indicating comparable perceptions of the IPPA across the two courses.

Taken together, these results suggest that students generally regarded the first in-person proctored assessment as a reasonable but imperfect measure of their prior mathematical knowledge. Compared with perceptions of the un-proctored summer placement test, the IPPA appears to have been viewed slightly more favorably, particularly among Calculus I students; however, the overall similarity of mean ratings across courses indicates that the shift to an in-person, proctored format did not dramatically alter students' perceptions of test accuracy (Figure 6).

Summary. Interestingly, comparisons between Survey Q1 and Survey Q2 indicate that students did not perceive the IPPA as significantly more accurate than the OUUPT in reflecting their precalculus skills. This perception persists despite evidence from prior work [3] showing that IPPA is a strong predictor of calculus performance, whereas OUUPT is not, as well as findings from our analyses of RQ1 and RQ2 demonstrating that approximately 50% of students' OUUPT scores fell outside the reasonable test–retest range (±25%). Several factors may help explain this discrepancy. Students may have interpreted the survey questions differently, leading to subjective or inconsistent judgments of "accuracy." For instance, some students may have used external resources when completing the OUUPT but responded to the survey assuming the assessment was taken without assistance. Under such assumptions, students may have viewed the OUUPT and IPPA—often considered parallel forms of the same assessment—as equally accurate. Additionally, students' self-assessments of prior knowledge may be imperfect, particularly early in the semester, limiting their ability to evaluate how well an assessment captures their true preparation. Other factors, such as familiarity with proctored, in-person exams, test anxiety, and the time gap between the OUUPT and IPPA, may have further influenced student perceptions. Collectively, these findings suggest that students' perceptions of placement accuracy do not necessarily align with the psychometric reliability or predictive validity of the assessments.

3.3.2. How do students perceive their placements under the current criteria?

<u>*Survey Q3.*</u> Before starting your first calculus math course at UVA, how confident were you that your placement was correct? ("1 = Very doubtful, ..., 5 = Very confident")

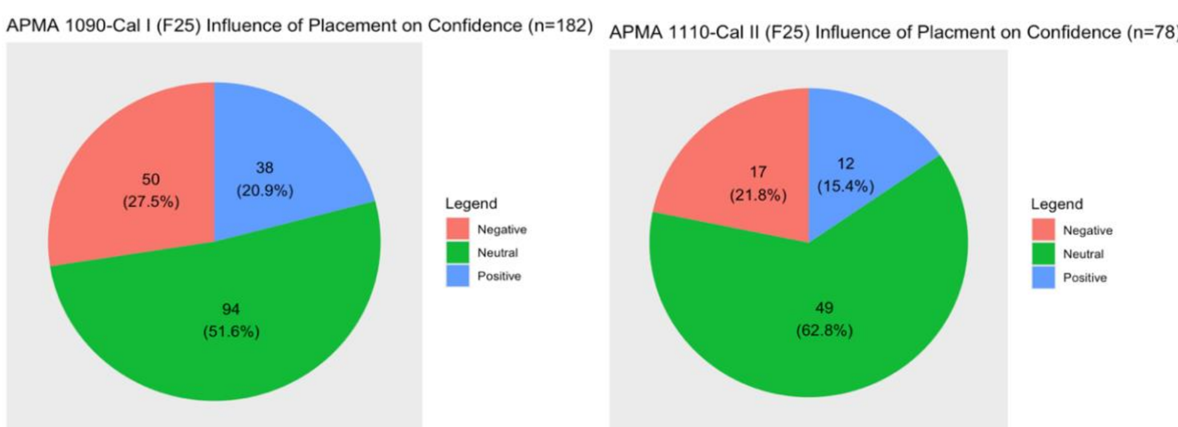


APMA 1090 mean=3.2 APMA 1110 mean=3.72

Figure 7: Students' Confidence in Placement

Students reported moderate-to-high confidence in the correctness of their placement, with markedly higher confidence among APMA 1110 students than APMA 1090 students. In APMA 1090 (n = 181), 45.3% reported high confidence (ratings 4–5), 26.5% were neutral, and 28.2% reported low confidence (1–2); the mean was 3.20 (95% CI: 3.03–3.38), with 13.8% "very confident" and 10.5% "very doubtful." In APMA 1110 (n = 78), confidence was higher: 61.5% rated high (including 35.9% "very confident"), 17.9% neutral, and 20.5% low; the mean was 3.72 (95% CI: 3.44–3.99). The between-course difference in means (Δ = 0.52, Cohen's $d \approx 0.42$) indicates a moderate advantage in pre-course confidence for Calculus II students. Overall, most students felt at least somewhat confident about their placement, but those placed into Calculus II expressed notably greater assurance than those in Calculus I (Figure 7).

<u>*Survey Q4.*</u> Did your placement influence your confidence in starting math at UVA? (-1 = Negative,0=Neutral, 3 = Positive)

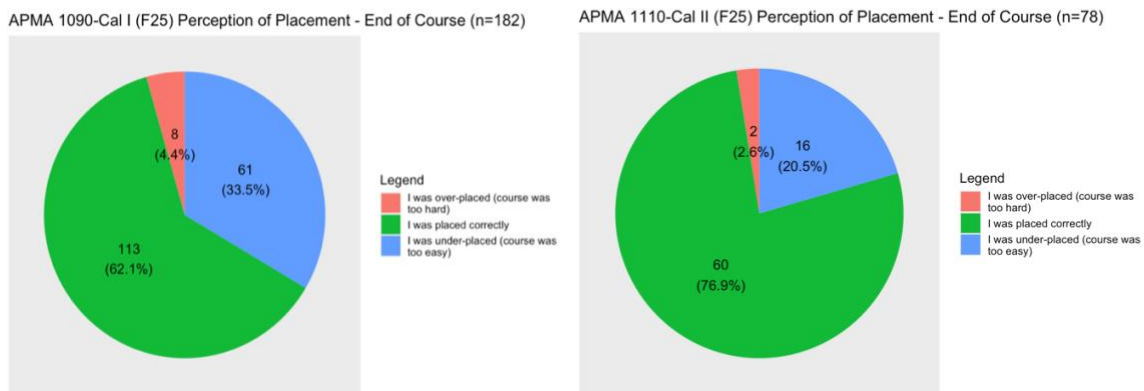


APMA1090 mean=-0.07 APMA 1110 mean=-0.06

Figure 8: Influence of Placement on Confidence

Across both courses, students most reported a neutral impact of placement on their confidence, followed by a positive impact. In APMA 1090, most students (51.6%) indicated that placement had a neutral effect, while 20.9% reported a positive effect and 27.5% reported a negative effect. Similarly, in APMA 1110, most students (62.8%) reported a neutral influence, with 15.4% indicating a positive effect and 21.8% a negative effect. Overall, these results suggest that placement decisions were unlikely to strongly influence students' initial confidence, particularly for those placed into Calculus II (Figure 8).

Survey Q5. After completing your first calculus math course at UVA, how do you feel about your placement? (1 = Under-placed, 0 = Correctly placed, -1 = Over-placed)



APMA1090 mean=0.29 APMA1110 mean=0.18

Figure 9: Perception of Placement

After completing their first calculus course, most students reported that they had been correctly placed, suggesting general alignment between placement decisions and students' academic experiences. Notably, In APMA 1090, 62.1% of students felt correctly placed, while 33.5% indicated that they were under-placed, meaning the course was too easy, and only 4.4% reported being over-placed, meaning the course was too hard. Similarly, in APMA 1110, 76.9% of students felt correctly placed, 20.5% felt under-placed (course too easy), and 2.6% felt over-placed (course too hard). Overall, perceptions of over-placement were rare in both courses, whereas a non-trivial proportion of students—particularly in Calculus I—felt that the course level did not sufficiently challenge them. These findings suggest that the current placement criteria may be slightly conservative, especially for students entering Calculus I (Figure 9).

Furthermore, we compared students' perceptions of the accuracy of their placement prior to entering the course (Survey Q3) with their perceptions after completing the course (Survey Q5) to determine whether any changes occurred (Figure 10).

For APMA 1090 (Calculus I), for students who ultimately reported being under-placed (course too easy), pre-course confidence was concentrated among the lower categories with a majority (65%) reporting being very doubtful or somewhat doubtful prior to the course and only 13% felt confident for their

placements. Students who later felt correctly placed exhibited a broader distribution of pre-course confidence, though most reported being neutral or somewhat confident, and relatively few reported being very doubtful. Among the small group who felt over-placed (course too hard), pre-course confidence was again spread across categories, with notable proportions reporting neutral or confident views before the course.

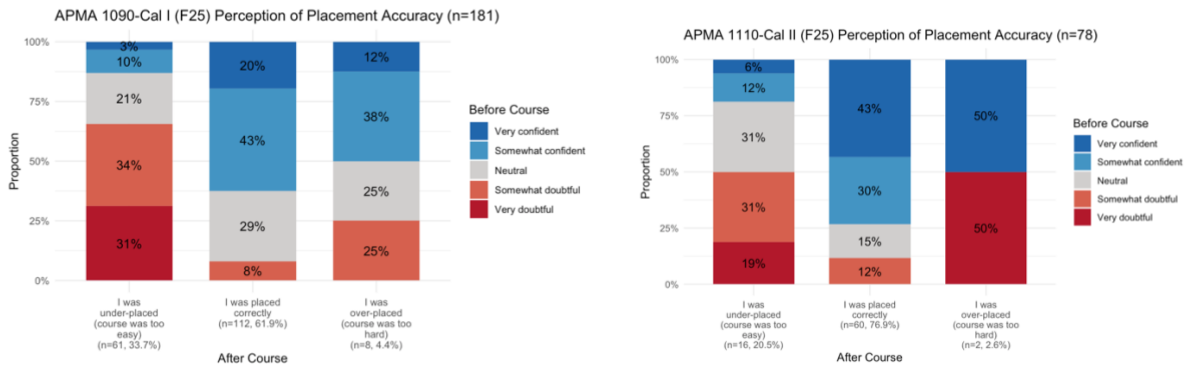


Figure 10: Pre-course placement confidence vs post-course placement perception

In APMA 1110, students who later reported being correctly placed primarily reported being somewhat or very confident prior to the course, though neutral and doubtful responses were also present. In the category of under-placed, like APMA 1090, only a small portion of students (18%) were confident that they were placed into the correct course.

Across both courses, Figure 10 indicates that students who ultimately perceived themselves as correctly placed generally expressed confidence prior to the course that their placement was appropriate. This pattern is particularly evident in APMA 1110, where there is substantial overlap across groups. In contrast, most students who later reported feeling either under-placed or over-placed had already expressed doubts about their placement before the course began. A Pearson correlation analysis comparing pre-course and post-course placement perceptions further supports this observation, revealing a moderately strong positive association ($r = 0.645$, $p < 0.001$). Together, these findings suggest that students' initial perceptions of placement accuracy are relatively stable over time and are meaningfully aligned with their post-course evaluations.

Survey Q6 and Q7 were open-ended questions and analyzed using an inductive thematic approach. Themes were identified from recurring patterns in student responses, and multiple themes could be assigned to a single response.

Survey Q6. What is your overall perception of the placement test? What improvements would you suggest for the placement test?

For Q6, two primary themes—perception and improvement—were established. Responses were coded into specific categories within these themes and summarized in Table 5 in Appendix 2. Most students viewed the placement test positively, indicating that it accurately reflected their mathematical readiness and placed them appropriately. For example, several students noted that the test helped them assess their preparedness for calculus. However, a notable portion of students reported misplacement, feeling that the test did not align with their prior preparation, often due to differences in high school curricula. Others expressed neutral views, stating that placement was primarily determined by transfer or AP credit, making the test less influential. Student suggestions for improvement focused mainly on:

- Timing and transparency, including clearer expectations and better preparation resources
- Content alignment, with more emphasis on calculus topics rather than precalculus
- Policy alignment, particularly better integration with AP/IB credit rules
- Flexibility, allowing adjustments to placement decisions

Survey Q7. If you did not perform well in the first proctored in-person precalculus assessment (IPPA) you took during your calculus class in F25, what are the possible factors?

For Q7, themes were derived inductively without predefined categories. The most common reasons students cited for underperformance on the first in-person assessment were:

- Content rustiness, due to time gaps since last studying precalculus
- Insufficient preparation, including unclear expectations or underestimating difficulty
- Stress and transition factors, especially anxiety and adjustment to college-level testing
- Assessment challenges, such as pacing or perceived difficulty

Less frequently, students mentioned gaps in prior knowledge or external personal factors. More details refer to Table 6 in Appendix 2.

Interviews. We recruited four students to conduct interviews, by using similar inductive approach as analyzing Survey Q6 & Q7, we found that Students generally viewed the math placement test as accurate and acknowledging that it correctly identified their skill levels even when they initially doubted the results. To improve the process, they suggested proctoring the summer test, so students treat it more seriously, shortening it to prevent students from "slacking off", and shifting the content to focus more on advanced calculus concepts rather than basic pre-calculus.

For those who struggled on the first in-person assessment, common failure factors included "rustiness" after the summer break, underestimating the test's difficulty, feeling pressured by tight time constraints, and not knowing that study resources were available beforehand.

The individual coding is attached in Appendix 3 at the end.

Summary. Under the current placement criteria, there is no significant impact on students' confidence to start math course after receiving their placement decisions. Most students felt at least somewhat confident that they were placed to the appropriate course 28.2% of Calculus I and 20.5% of Calculus students were skeptical about their placement prior to the course. After completing the course, 37.9% Calculus I students and 23.1% Calculus II students perceived that their placement was not correct with dominant of those students perceived that the course was too easy. In addition, students who believed that their course was not appropriate for them after completing it already expressed doubts about their placements prior to the start of the course, but they were still required to take the course since they did not have the credit of the course. Taken together, these findings raise the possibility that the current placement criteria may be slightly conservative, particularly with respect to Calculus I placement decisions. We may need to revise the current criteria for furfure student to benefit students the most.

Meanwhile, the qualitative analysis revealed that most students generally viewed the placement test as an accurate indicator of their mathematical readiness, though a notable subset reported misalignment with their actual preparation or limited impact due to credit-based placement policies. Students suggested improvements related to clearer expectations, better timing, increased alignment with calculus content, and greater transparency about how test results are used. Underperformance on IPPA was most often attributed to content rustiness, insufficient preparation, stress related to the college transition, and concerns about assessment design.

3.3.3. *Whether differences exist in OUPPT, IPPA, and final exam scores according to students' perceptions of their placement after completing the course*

It is noteworthy that a nontrivial proportion of students reported feeling under-placed (33.5% in Calculus I and 20.4% in Calculus II). This observation motivated further investigation into which students held this perception after completing the course. Based on students' self-reported perceptions, we categorized them into three groups—over-placed, correctly placed, and under-placed—and examined whether differences existed in their OUPPT, IPPA, and final exam scores.

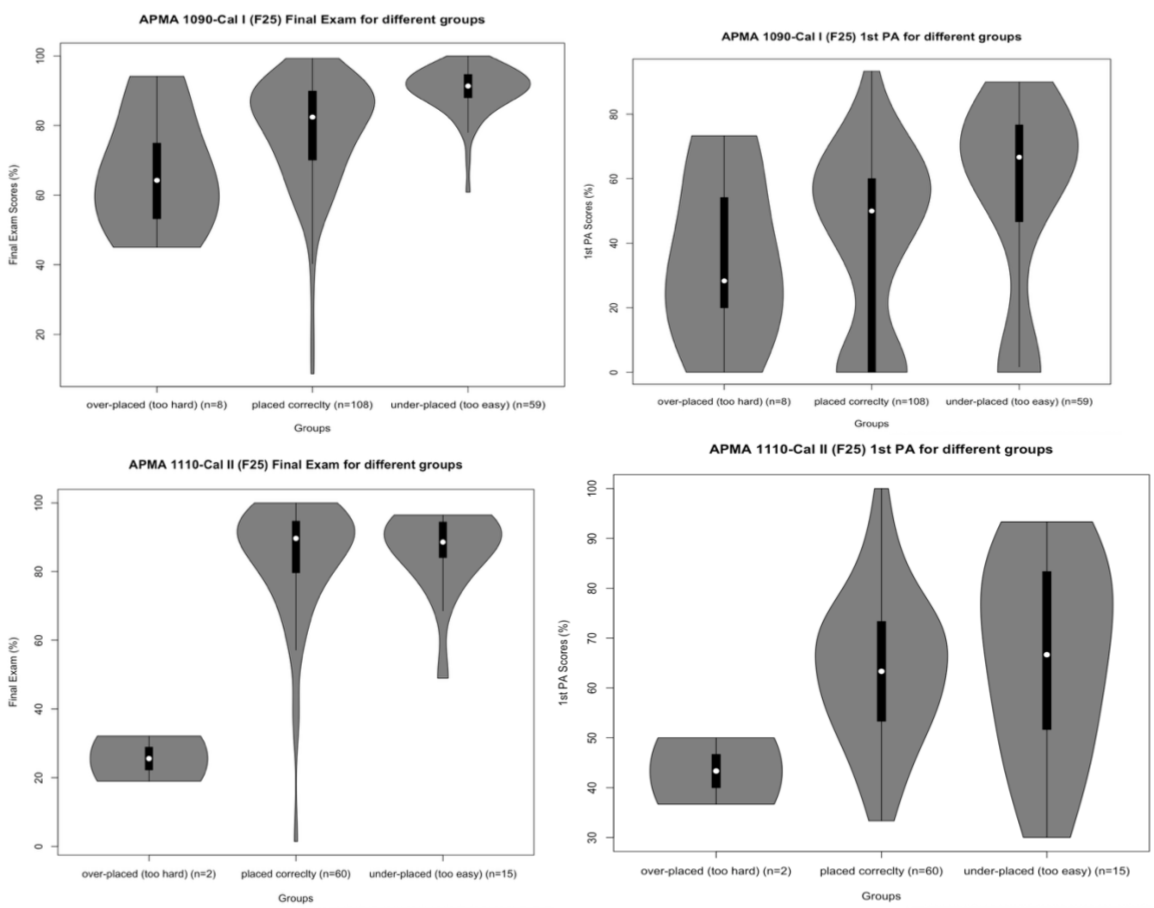


Figure 11: IPPA and Final Exam Violin Plots for Different Groups

Given concerns about the reliability of the OUPPT, we focus our analysis on IPPA and final exam performance across placement perception groups. The violin plots (Figure 11) suggest clear performance differences aligned with students' self-reported placement perceptions. In both Calculus I and Calculus II, students who perceived themselves as under-placed generally achieved higher median IPPA and final exam scores, with distributions shifted upward relative to the correctly placed group. In contrast, over-placed students tended to exhibit substantially lower IPPA and final exam performance, particularly pronounced in Calculus II, where their score distributions were concentrated at the lower end. The correctly placed group typically fell between these two extremes, both in central tendency and variability. These patterns indicate that students' retrospective perceptions of placement are meaningfully associated with objective measures of course performance. Overall, the results suggest that student self-perception provides useful diagnostic information and may complement existing placement tools when identifying misalignment between course level and student preparedness.

When analyzing differences in students' OUPPT and IPPA by their perceptions of the accuracy of their placement by ANOVA method, we observed significant differences between Calc I students who perceived their placement to be too easy compared to those who perceived their placement as accurate. In particular, students who reported the course placement as too easy performed significantly higher on the OUPPT and IPPA compared to students who reported the placement as accurate, $F_{OUPPT}(1,157) = 10.263$, $p=.002$, $F_{IPPA}(1,165) = 8.190$, $p=.005$. These differences did not exist between Calc II student who perceived their placement as too easy compared to those who perceived their placement as accurate. We observed similar trends in exam scores based on student perceptions of placement accuracy. Specifically, in Calc I we observed students who perceived the placement as too easy performed significantly higher on the final exam (M=90.15, SD=6.88) than students who perceived the placement as accurate (M=78.16, SD=15.85), $F(1,165) = 30.542$, $p=.000$. No significant differences were observed in Calc II final exam scores between students who perceived the placement as accurate (83.83, SD=17.36) compared to students who perceived the placement as too easy (86.27, SD=12.11).

	Calc I: Placed Correctly (n=102)			**Calc I: Under-placed (Too easy (n=57)**		
	OUPPT	**IPPA**	**Final**	**OUPPT**	**IPPA**	**Final**
OUPPT	--			--		
IPPA	-.078	--		.045	--	
Final	.141	.274*	--	-.116	.232	--
	Calc II: Placed Correctly (n=55)			**Calc II: Under-placed (Too easy (n=16)**		
	OUPPT	**IPPA**	**Final**	**OUPPT**	**IPPA**	**Final**
OUPPT	--			--		
IPPA	.413*	--		.482	--	
Final	.213	.410*	--	-.316	.527*	--

Table 6: Pearson Correlation Coefficients for groups (Note. * $p<.05$)

When examining the relationship between student's placement perceptions, OUPPT, IPPA, and final exam scores by using correlation tests, there were significant relationships between IPPA and final exam scores for some students (Table 6). For all students except those in Calc I who perceived their placement as too easy, the IPPA was significantly and positively correlated with the final exam grade. We also observed that there were no relationships between OUPPT and IPPA performance for Calc I students, regardless of their placement perceptions. In Calc II, there were strong positive correlations between students' OUPPT and IPPA. Interestingly, for students in either Calc I or Calc II who perceived their placement as too easy, their OUPPT was negatively correlated with final exam scores.

Summary. Overall, regardless of students' post-course perceptions of their placement, IPPA scores were positively associated with final exam performance, whereas OUUPT scores were not. This result further reinforces the findings in [3] that IPPA is an important predictor of calculus performance, while OUUPT is not. Among Calculus I students, those who perceived themselves as under-placed after completing the course performed significantly better on both the IPPA and the final exam than their peers, a pattern that was not observed in Calculus II. In contrast, students who perceived themselves as over-placed after course completion experienced significantly greater difficulties on the IPPA and final exam, regardless of whether they were enrolled in Calculus I or Calculus II. Together, these findings suggest that the current placement criteria, particularly for Calculus I, may be somewhat conservative.

4. Conclusion and Discussion

This study examined differences between an online un-proctored precalculus placement test (OUPPT) and an in-person proctored precalculus assessment (IPPA), with particular attention to readiness for Calculus I and Calculus II, domain-level consistency across assessments, and students' perceptions of their placements. Collectively, the results suggest that while the two assessment modalities provide broadly aligned signals for students with strong precalculus preparation, meaningful discrepancies persist for a substantial subset of students, particularly those performing near key placement thresholds. These findings highlight important considerations for the design, interpretation, and implementation of placement assessments in undergraduate mathematics.

OUPPT and IPPA yield relatively consistent readiness indicators for students with strong precalculus preparation; however, substantial discrepancies beyond typical test–retest variation (±25%) are common among lower- and mid-performing students, representing approximately half of the sample. These discrepancies are especially pronounced for students placed into Calculus II, indicating increased uncertainty near critical placement thresholds. Contributing factors likely include differences in testing conditions, the un-proctored nature of the OUPPT, and less stable mastery of precalculus content among lower-performing students.

Several factors may contribute to these modality-based discrepancies. The un-proctored nature of the OUPPT may permit differential access to external resources or assistance [23], [24], potentially inflating scores for students whose precalculus knowledge is less secure. Differences in testing environments and perceived stakes may also influence student performance [25]. In-person assessments, particularly early in the semester, may be affected by test anxiety, adjustment to college-level expectations, or unfamiliar pacing demands. At the same time, greater score variability among lower- and mid-performing students may reflect less stable mastery of foundational content, increasing susceptibility to contextual and cognitive factors across testing conditions.

Domain-level analyses further clarify the sources of score divergence between modalities. Algebra-focused learning targets—including symbolic manipulation, exponents and logarithms, equation solving, and absolute value inequalities—consistently exhibited the largest discrepancies between OUPPT and IPPA. Learning targets related to limits and continuity showed similar patterns, with lower accuracy on the IPPA and larger modality gaps. In contrast, trigonometric knowledge and function graphing tasks demonstrated higher accuracy and greater consistency across assessments. These patterns were present for both Calculus I and Calculus II students but were more pronounced among those entering Calculus II, where OUPPT scores more frequently exceeded IPPA performance on algebraic and limits-based topics.

Taken together, these findings indicate that agreement between online and in-person assessments is not uniform across mathematical domains. Instead, modality alignment varies systematically, suggesting that discrepancies in overall placement scores may mask meaningful differences in the types of knowledge being assessed. Algebraic manipulation and limits, which are critical for early success in calculus, appear particularly sensitive to assessment context and may warrant additional scrutiny when used for placement purposes. From a practical standpoint, these results suggest value in supplementing overall placement scores with domain-level information, especially when placement decisions depend on performance in these foundational areas.

Students' perceptions of the placement process provide additional context for interpreting these quantitative results. Under the current placement criteria, most students reported feeling at least somewhat confident in their placement decisions, suggesting that the process is generally viewed as credible and reasonable. However, a notable minority expressed doubts about their placement prior to beginning their course, and this skepticism increased after course completion, particularly among Calculus I students. After completing the course, many of these students felt that their placement had been inappropriate, most commonly reporting that the course was too easy. Importantly, students who ultimately believed their placement was incorrect often had expressed similar concerns beforehand but lacked the flexibility to adjust their placement due to credit requirements or institutional policies.

Qualitative data from surveys and interviews shed further light on these perceptions. While many students described the placement test as an accurate indicator of their mathematical readiness, others pointed to misalignment between test content and their prior coursework or mathematical strengths. Students frequently cited timing and transparency as areas for improvement, expressing a desire for clearer communication about test expectations, greater access to preparation materials, and placement testing closer to the start of the term. Several students also questioned the value of the placement test when credit-based policies limited their ability to act on the results. In addition, underperformance on the IPPA was commonly attributed to content rustiness over the summer, insufficient preparation, stress associated with the transition to college, and aspects of assessment design such as pacing and length.

These findings have several implications for placement practice. First, the results suggest that a single assessment modality, especially online un-proctored, may be insufficient to support high-stakes placement decisions for students near placement cutoffs. Incorporating targeted follow-up assessments or domain-specific verification, particularly for algebraic and limits-based skills, may help reduce misplacement risk. Second, increased transparency about assessment content, timing, and the role of placement scores relative to credit policies may improve student trust and engagement with the process. Finally, early diagnostic assessments and support mechanisms within Calculus I and II could help identify and address gaps in prerequisite knowledge, mitigating the consequences of imperfect placement decisions.

This study is subject to several limitations. The findings reflect data from a limited number of cohorts at a single institution, which may constrain generalizability to other contexts with different curricula, student populations, or placement policies. In addition, differences in assessment format, timing, and stakes complicate direct score comparability between OUPPT and IPPA. Students' self-reported perceptions, while informative, may also be influenced by retrospective bias or course outcomes. Future research could explore experimental or quasi-experimental designs to better isolate modality effects, examine longitudinal outcomes associated with different placement pathways, and investigate how domain-level diagnostic information can be most effectively integrated into placement systems.

Overall, this study underscores the complexity of assessing precalculus readiness and making accurate placement decisions in a large-scale instructional setting. While online un-proctored assessments offer accessibility and efficiency, their alignment with in-person measures varies meaningfully across student groups and mathematical domains. A more nuanced, domain-aware approach to placement—combined with clearer communication and flexible support structures—may better serve students and promote successful entry into the calculus sequence.

5. Acknowledgments

This research received approval from the Institutional Review Board (IRB) at the University of Virginia. All study procedures and data handling practices adhered strictly to the approved IRB standards. We sincerely appreciate the faculty and instructors of APMA 1090 and APMA 1110 for generously sharing course performance data that made this analysis possible. We also extend our thanks to the reviewers for their insightful feedback, which contributed significantly to improving the clarity and rigor of this paper.

Reference

[1] M. Hodara, S. Smith, J. Melinda, and M. Karp, “Improving Developmental Education Assessment and Placement: Lessons From Community Colleges Across the Country,” 2012.

[2] Judith Scott-Clayton, “Evidence-based reforms in college remediation are gaining steam – and so far living up to the hype.” Accessed: Jan. 17, 2026. [Online]. Available: https://www.brookings.edu/articles/evidence-based-reforms-in-college-remediation-are-gaining-steam-and-so-far-living-up-to-the-hype/

[3] M. Li, “Predictive Validity of Online and In-Person Precalculus Placement Tests in Calculus I and II”.

[4] M. Li, S. Pisano, J. F. Marley, A. M. Fernando, and L. Wheeler, “Integrating Precalculus into Calculus II and Its Outcomes,” in *ASEE Annual Conference and Exposition, Conference Proceedings*, 2024. doi: 10.18260/1-2--47659.

[5] M. Li, S. Pisano, and UVA APMA Faculty, “UVA APMA PreCalculus-Preparation.” [Online]. Available: https://apmauva.github.io/PreCalculus-Preparation/

[6] J. Scott-Clayton, “Do High-Stakes Placement Exams Predict College Success?,” 2012.

[7] R. George, “Predictors for Success in Calculus I,” 2020. doi: 10.18260/1-2--33184.

[8] O. Ryan, S. Sajadi, S. Barrera, and R. T. Jaghargh, “Understanding the Effects of a Math Placement Exam on Calculus 1 Enrollment and Engineering Persistence,” *Educ. Sci. (Basel).*, vol. 15, no. 2, Feb. 2025, doi: 10.3390/educsci15020154.

[9] D. Bullock, J. Callahan, Y. Ban, A. Ahlgren, and C. Schrader, “The implementation of an online mathematics placement exam and its effects on student success in precalculus and calculus,” in *ASEE Annual Conference and Exposition, Conference Proceedings*, 2009. doi: 10.18260/1-2--5541.

[10] C. Newhouse, J. Squire, and M. Swenty, “Math Placement Testing: Performance and Persistence in Civil Engineering,” 2025. doi: 10.18260/1-2--54211.

[11] S. Varol and Z. Ridder, “Analysis of Engineering Student Retention Based on Math Placement and Performance,” *Journal of College Orientation, Transition, and Retention*, vol. 31, no. 2, 2024, doi: 10.24926/jcotr.v31i2.5978.

[12] P. A. Westrick, J. P. Marini, and E. J. Shaw, “An In-Depth Exploration of SAT ® Math Scores for Use in College Course Placement.”

[13] “AP placement validity study,” College Board. Accessed: Jan. 17, 2026. [Online]. Available: https://aces.collegeboard.org/aces-studies/ap-placement-validity-study

[14] “ALEKS placement, preparation, and learning: Research overview.” Accessed: Jan. 17, 2026. [Online]. Available: https://www.seattleu.edu/science-engineering/academic-departments/department-of-mathematics/math-placement/

[15] M. Ferrer, C. Rueda, and J. Morales, “Adaptive online mathematics placement testing in engineering education,” *Comput. Educ.*, vol. 142, 2019.

[16] A. J. Bowers and X. Zhou, “Learning analytics and online mathematics placement exams: Predicting calculus success for engineering students,” *Journal of Learning Analytics*, vol. 8, no. 2, pp. 45–62, 2021.

[17] Stefani Relles and William G. Tierney, “Equity Implications of Online Placement Testing in STEM Pathways,” *Journal of Higher Education*, vol. 93, no. 4, pp. 589–615, 2022.

[18] K. Miller and R. Best, “Validity of remote online math placement testing for engineering students,” *International Journal of STEM Education*, vol. 10, no. 37, 2023.

[19] S. Naqvi, R. Srivastava, T. Al Damen, A. Al Aufi, A. Al Amri, and S. Al Adawi, “Establishing Reliability and Validity of an Online Placement Test in an Omani Higher Education Institution,” *Languages*, vol. 8, no. 1, 2023, doi: 10.3390/languages8010061.

[20] Ronald Jay Cohen and Mark E. Swerdlik, *Psychological Testing and Assessment: An Introduction to Tests and Measurement*, 10th ed. New York.: McGraw Hill LLC, 2022.

[21] C. B. Terwee *et al.*, “Quality criteria were proposed for measurement properties of health status questionnaires,” *J. Clin. Epidemiol.*, vol. 60, no. 1, 2007, doi: 10.1016/j.jclinepi.2006.03.012.

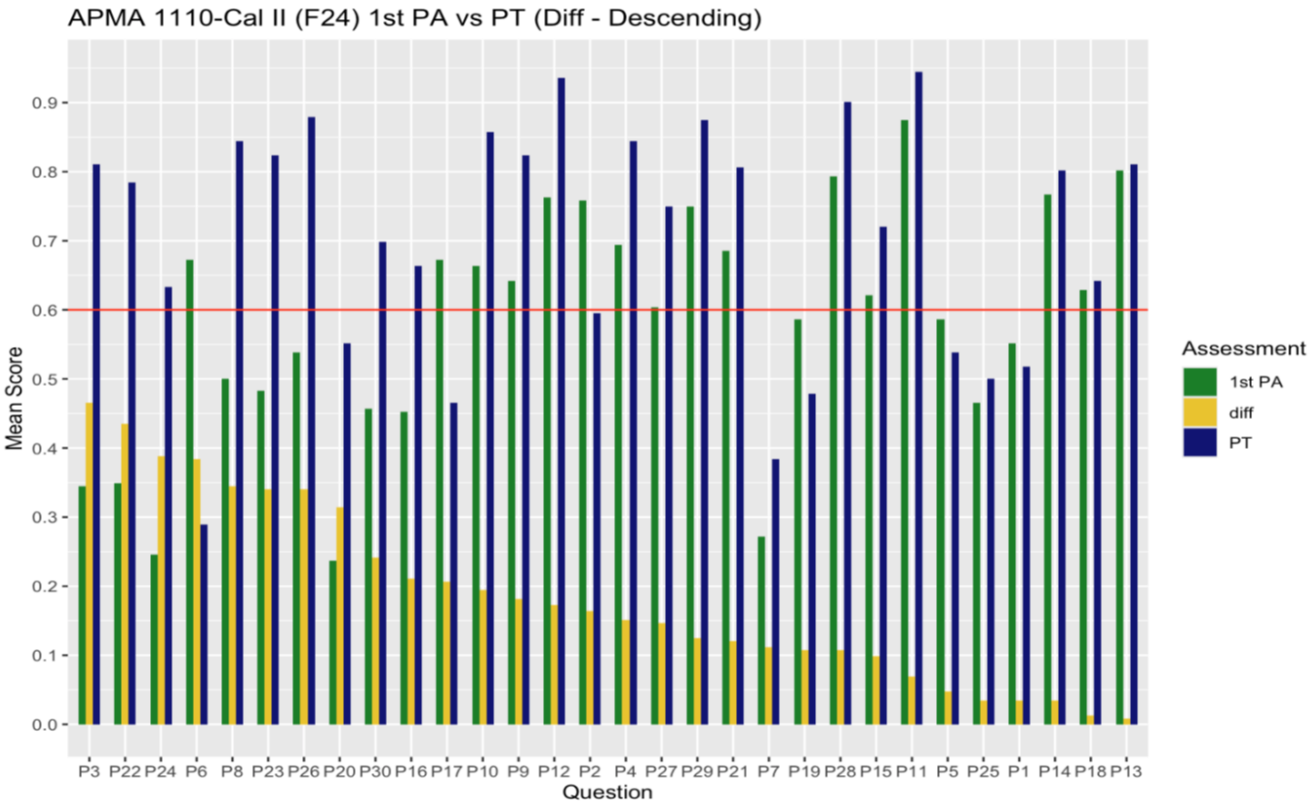
[22] P. Irwing, T. Booth, and D. J. Hughes, *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*, vol. 1–2. 2017. doi: 10.1002/9781118489772.

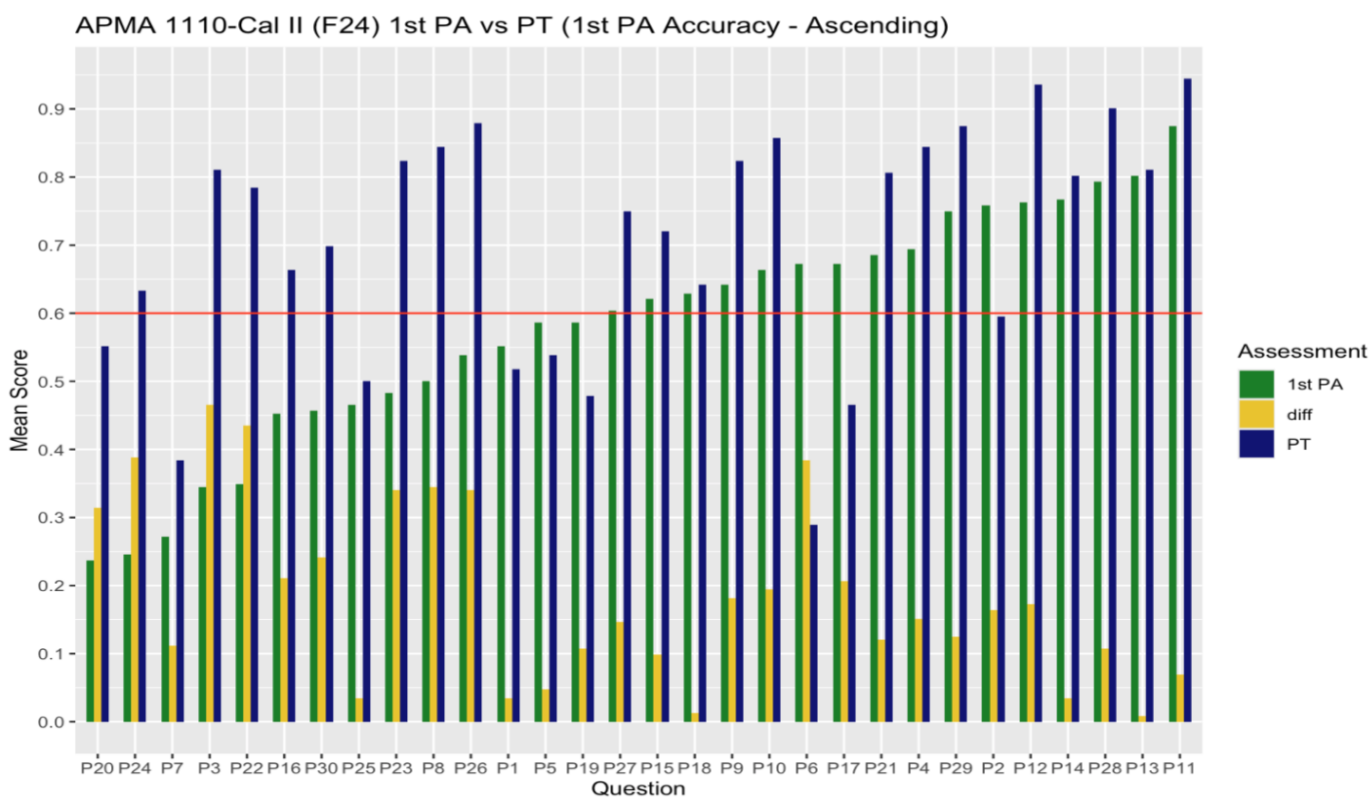
[23] J. Pleasants, J. M. Pleasants, and B. P. Pleasants, “Cheating on Unproctored Online Exams: Prevalence, Mitigation Measures, and Effects on Exam Performance,” *Online Learning Journal*, vol. 26, no. 1, 2022, doi: 10.24059/olj.v26i1.2620.

[24] P. M. Newton and K. Essex, “How Common is Cheating in Online Exams and did it Increase During the COVID-19 Pandemic? A Systematic Review,” 2024. doi: 10.1007/s10805-023-09485-5.

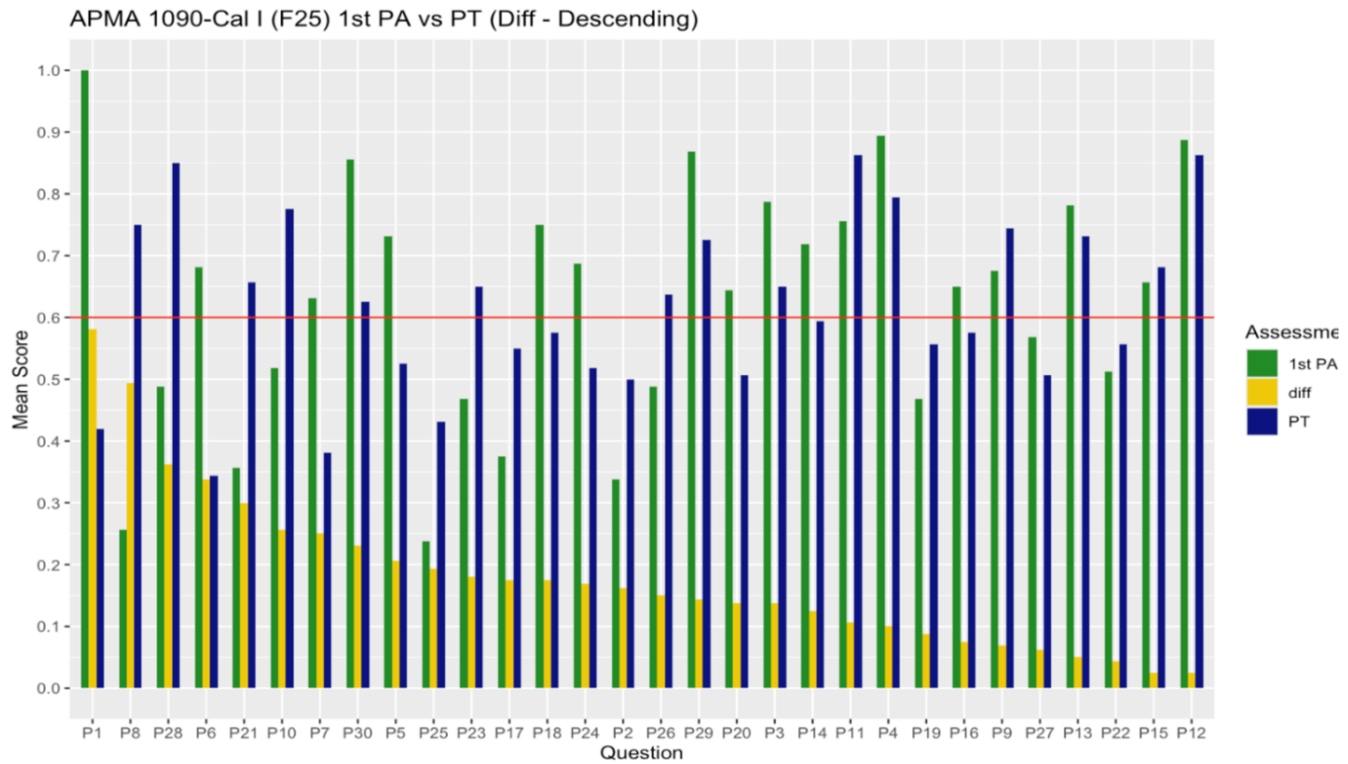
[25] S. Caviola, E. Toffalini, D. Giofrè, J. M. Ruiz, D. Szűcs, and I. C. Mammarella, “Math Performance and Academic Anxiety Forms, from Sociodemographic to Cognitive Aspects: a Meta-analysis on 906,311 Participants,” *Educ. Psychol. Rev.*, vol. 34, no. 1, 2022, doi: 10.1007/s10648-021-09618-5.

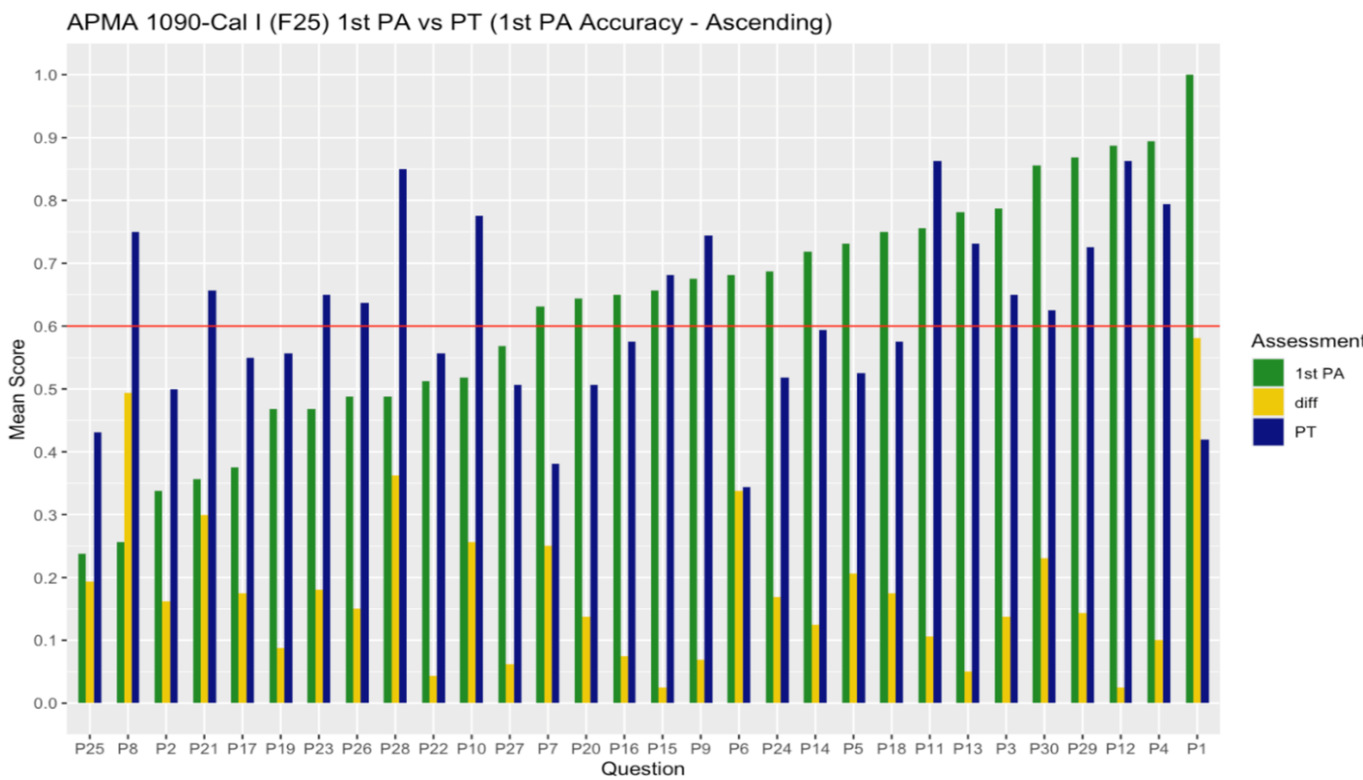
Appendix 1


APMA 1110-Cal II (F24) 1st PA vs PT (Diff - Descending)
Mean Score
Question
Assessment
1st PA
diff
PT
P3 P22 P24 P6 P8 P23 P26 P20 P30 P16 P17 P10 P9 P12 P2 P4 P27 P29 P21 P7 P19 P28 P15 P11 P5 P25 P1 P14 P18 P13

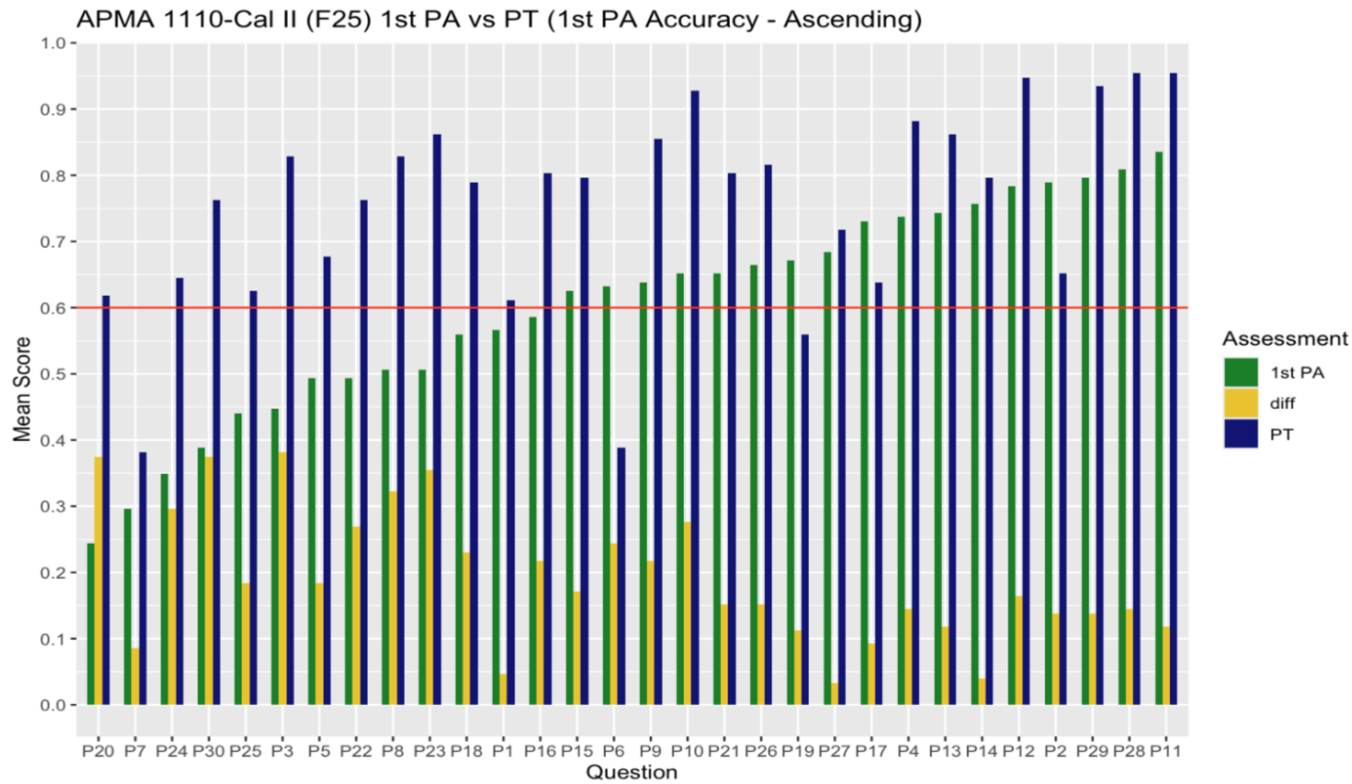

APMA 1110-Cal II (F24) 1st PA vs PT (1st PA Accuracy - Ascending)
Mean Score
Question
Assessment
1st PA
diff
PT
P20 P24 P7 P3 P22 P16 P30 P25 P23 P8 P26 P1 P5 P19 P27 P15 P18 P9 P10 P6 P17 P21 P4 P29 P2 P12 P14 P28 P13 P11

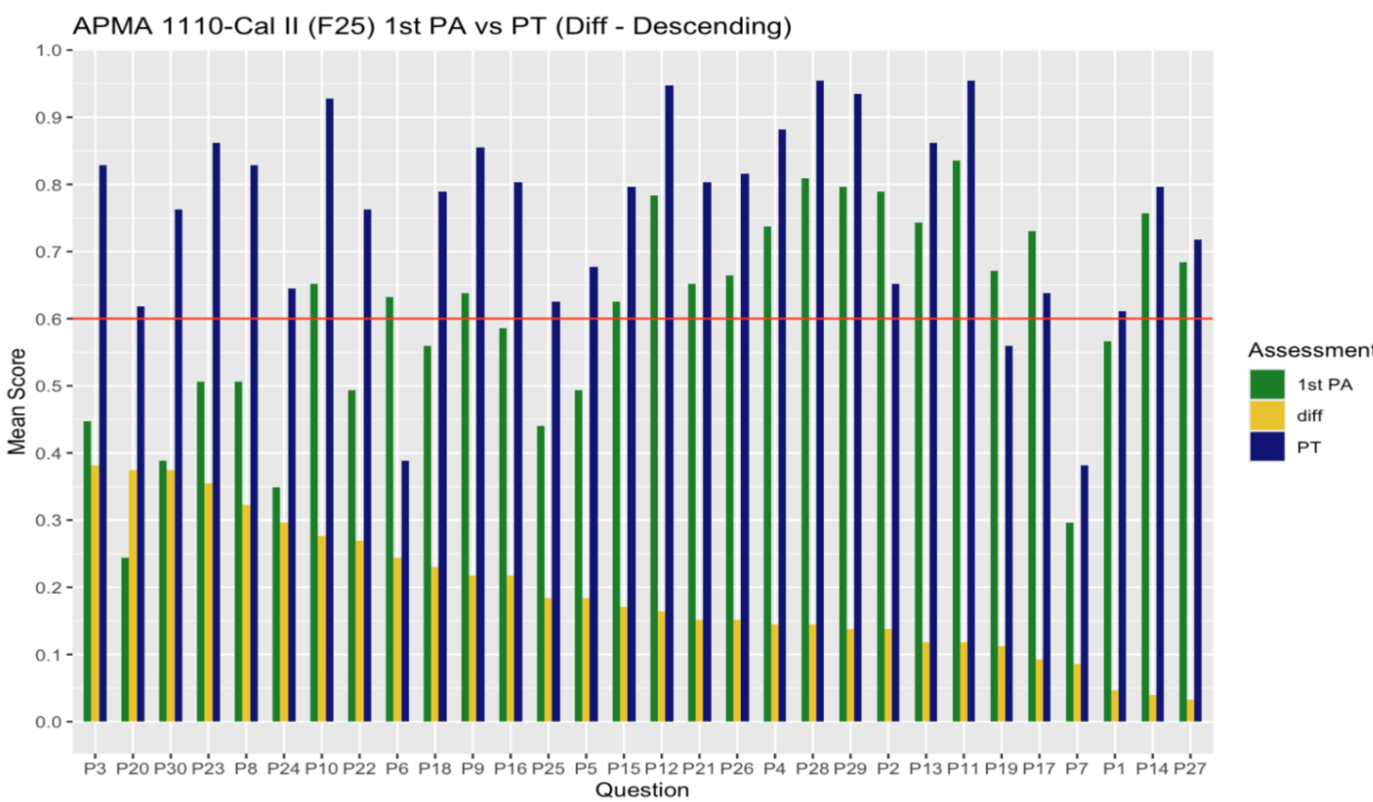
Appendix 1-Cont.





Appendix 1-Cont.





Appendix 1-Cont.

Fall 2024 APMA 1110 (Calculus II) cohort. For the Fall 2024 APMA 1110 (Calculus II) cohort, bar plots reveal patterns that are broadly consistent with those observed for Calculus I but with more pronounced discrepancies between OUPPT and IPPA. The largest differences again occur primarily in algebra-focused learning targets—including simplifying expressions, laws of exponents and logarithms, solving equations and inequalities, and absolute-value problems—as well as in topics related to limits and continuity. In contrast, trigonometric and function-graphing learning targets (e.g., unit circle, trigonometric identities and graphs, and graphing of power, exponential, and logarithmic functions) display relatively smaller differences and higher accuracy on both assessments. Compared with Calculus I, Calculus II students exhibit a more consistent tendency for OUPPT scores to exceed IPPA scores across algebra-related topics, suggesting greater divergence between assessment modalities at this course level.

Fall 2025 APMA 1090 (Calculus I) cohort. For the Fall 2025 APMA 1090 (Calculus I) cohort, the bar plots show topic-level patterns similar to those observed in Fall 2024 Calculus I, with discrepancies between OUPPT and IPPA varying systematically by learning target. Items associated with algebraic manipulation and equation solving—including simplifying expressions, laws of exponents and logarithms, polynomial division, and solving equations or inequalities—appear most frequently among those with larger absolute differences. As in prior cohorts, these items often show higher mean scores on OUPPT than on IPPA, indicating greater divergence for foundational algebra skills. In contrast, trigonometry- and graphing-related learning targets—such as evaluation of trigonometric values, use of the unit circle, trigonometric graphs, and graphing of power, exponential, and logarithmic functions—exhibit relatively smaller differences and higher IPPA accuracy. Items linked to limits and continuity again fall toward the lower end of IPPA accuracy and display moderate to larger discrepancies. Overall, the Fall 2025 Calculus I results reinforce the pattern that topic-level alignment between OUPPT and IPPA is stronger for trigonometric and graphical concepts and weaker for algebra-focused and limits-based topics, consistent with findings from other Calculus I cohorts.

Fall 2025 APMA 1110 (Calculus II) cohort. For the Fall 2025 APMA 1110 (Calculus II) cohort, the bar plots show a pattern consistent with other Calculus II cohorts, characterized by larger and more systematic discrepancies between OUPPT and IPPA than those observed in Calculus I. Items exhibiting the greatest differences are predominantly associated with algebra-focused learning targets**,** including simplifying expressions, laws of exponents and logarithms, polynomial division, and solving equations and inequalities, as well as with limits-related topics**.** For many of these items, OUPPT mean scores exceed IPPA mean scores by a noticeable margin. In contrast, items linked to trigonometric concepts and function-graphing learning targets—such as the unit circle, trigonometric identities and graphs, and graphing power, exponential, and logarithmic functions—demonstrate higher IPPA accuracy and relatively smaller OUPPT–IPPA differences. Overall, the Fall 2025 Calculus II results reinforce the trend that topic-level discrepancies between online and in-person assessments are more pronounced for algebraic and limits-based domains, while trigonometric and graphical topics remain comparatively stable across testing modalities.

Appendix 2

N=269	**Theme Labels**	**Theme Descriptions**	**Counts**	**Sample response**
Perception Themes	Accurate Placement	Placement test accurately reflected the student's math readiness or placed them appropriately.	60	"put me in the right place despite having taken Calculus BC last year,*"* *"helpful to gauge where I was in the course."*
	Misplacement	Placement test did not reflect ability and placed the student too high or too low.	25	*"did not reflect my knowledge of math and placed me too low.*"
	Neutral Perception	Placement test had little effect or elicited no strong opinion.	22	*"The placement test didn't really determine what course I was placed in," "wasn't necessary."*
	Stressful	Placement test was stressful or intimidating regardless of outcome.	2	*"it was stressful to take the placement test"*
Improvement Themes	Content Adjustment	more calc, less pre-calc, broader coverage	21	*"reflect more of the Calculus 1 material we actually see in class."*
	Test Design Changes	test format, length, difficulty level, show work, time pressure	8	*"more time to work on the problem"*
	Timing and Transparency	improve timing, preparation resources, expectations or communication	23	*"sending students an email about what could be on the placement test"*
	Credit Policy alignment	Better align placement test with credit policies, a 4 on the AP exam should give credit	17	*"I don't see the point if I do well on the test but still can't move up because I don't have credit."*
	Placement Flexibility	allow overrides and more flexibility to move up or down	10	*"loose the AP requirement to place students"*

Table 5

Appendix 2-Cont.

Theme Labels	**Theme Descriptions**	**Counts**	**Sample response**
Content Rustiness	Long time gap since taken precalculus	43	*"It had been a while since I learned the topic presented,"*
Insufficient Preparation	Insufficient study time/lack of preparation, unclear preparation expectation	42	*"I was not very well informed in what would be on the test,"* *"I under-studied because I had already completed a calculus course, so I thought I already knew the material."*
Gaps in Prior Knowledge	missing foundational content dues to prior coursework differences, Never learned certain material	18	*"the test had material that was far more advanced than anything I'd seen in pre-calc,"*
Assessment Design Challenges	Misalignment in test content, pacing or structure, questions are hard	19	*"time was too short to complete some challenging problems"*
Stress and Transition Factors	Anxiety or difficulty adjusting to college-level testing	25	*"Anxiety of my first in-person assessment at UVA,"* *"stress, tired, adjusting to new college life."*
Personal Circumstances	External personal factors	3	
Non-response	Non-response/Not Applicable (mastery, I passed, I don't remember)	17	*"I don't remember"*

Table 6

Appendix 3

Student Interview 1

Perception Themes

1. Accurate Placement: Student_1 felt the summer test correctly judged his math skills because it covered topics he either knew or didn't know.

   Quote: "I think it was pretty accurate".

2. Neutral Perception: He did not put in much effort because he felt the test was just a guide and he had already decided which class to take.

   Quote: "I didn't take it very seriously".

3. Stressful: While the summer test felt relaxed, he found the in-person assessment during the semester difficult due to the speed required.

   Quote: "I think the time constraint was pretty hard".

Improvement Themes

1. Test Design Changes: Student_1 suggested that shorter tests would help students stay focused and provide better results.

   Quote: "It would be better on the mind if it was shorter... they might be slacking off on the test if it's super long".

2. Timing and Transparency: He struggled with the in-person assessment because he hadn't practiced math recently.

   Quote: "I was very rusty on precalculus".

3. Placement Flexibility: Even though the test recommended a higher class, he chose to repeat an easier one to feel more prepared.

   Quote: "I took Calculus 1 again to be extra sure I knew everything".

Factors for Poor Performance:

1. Content Rustiness: "I was very rusty on precalculus".

2. Assessment Design Challenges: "I thought it was very fast paced".

3. Insufficient Preparation: He admitted he did not prepare for the first attempt.

Appendix 3-Cont.

Student Interview 2

Perception Themes

1. Accurate Placement: Before the class began, he believed he was placed too low because of his high school background. Student_2 ultimately felt the test correctly placed him because the class level felt right once he started.

   Quote: "At first, I didn't think it was correct... I thought [my AP scores] were enough for me to place into Calculus 2."

   Quote: "It was accurate for me. I felt perfectly placed in Calculus 1."

Improvement Themes

1. Test Design Changes (Proctoring): Student_2 strongly suggested that the test should be supervised to stop students from cheating to skip classes.

   Quote: "Maybe it should be proctored... students would treat it more like a real test."

2. Credit Policy Alignment: He felt that his previous high school success should have been enough to move him forward.

   Quote: "I took Calculus AB and BC in high school... I thought it was enough."

3. Timing and Transparency: He noted that students don't study during the summer, which makes the results less reliable.

   Quote: "Students aren't really studying in the summer... so it might not fully reflect a student's performance."

Factors for Poor Performance:

1. Insufficient Preparation: "In my mind, I thought it was gonna be very, very easy," so he didn't prepare.

Appendix 3-Cont.

Student Interview 3

Perception Themes

1. Accurate Placement: Student_3 felt the summer placement test was very accurate, giving it a score of 4 out of 5. She was placed in Calculus 2 and found it to be the perfect challenge level.

   Quote: "I definitely think, yeah, it was perfect. I really think it, like, met me where my level was at".

2. Neutral Perception: After seeing her result, Student_3's attitude did not change.

   Quote: "I didn't really feel anything, to be honest... It didn't really affect my perception of how I would do".

Improvement Themes

1. Content Adjustment: Student_3 suggested that since engineering students should already know basic math, the test should focus more on harder calculus topics.

   Quote: "Maybe less of a pre-calc focus, and maybe more of a focus on... Calculus 1, 2, or 3".

2. Test Design Changes: She felt that having only 15 questions per section didn't give a complete view of what a student actually understands.

   Quote: "It's just not a good full picture... because you only get 15 questions, it's just not a good full picture demonstration".

3. Timing and Transparency: Student_3 noted that the summer test is a poor predictor of success because students don't study for it like they do once classes actually start.

   Quote: "There's just a lot of things you're gonna forget over the summer... a little more time to prepare would be better".

Factors for Poor Performance:

1. Non-response (Passed): "Yeah, I passed it"

Appendix 3-Cont.

Student Interview 4

Perception Themes

1. Accurate Placement: Although she initially disagreed, Student_4 realized after taking the class that the lower placement was actually the right choice for her.

   Quote: "I felt like I could be placed into Calc 2, so I thought that it was incorrect".

   Quote: "I think after Experience Calc 1, it was correct, because I struggled a lot".

2. Stressful: Student_4 finds testing very difficult and often feels overwhelmed or "blanks out" during exams.

   Quote: "I'm a really bad test taker... I just, like, go to the test, and I'm just, like, completely blank".

Improvement Themes

1. Content Adjustment: She suggested that tests should measure how well a student understands a concept, rather than just whether they can memorize a formula.

   Quote: "Actually understanding, like, the concepts... is a completely different thing, so it definitely is hard for that sort of knowledge to be seen on a standardized test".

2. Timing and Transparency: Student_4 mentioned that she was not aware of the study tools available to her until it was too late.

   Quote: "I didn't know about them [the resources] until after we took them".

Factors for Poor Performance:

1. Stress and Transition: "I'm a really bad test taker... I just, like, go to the test, and I'm just, like, completely blank".

2. Gaps in Prior Knowledge: "I've, like, never really mastered that type of concept".

3. Insufficient Preparation: "I didn't know about them [the resources] until after we took them".