Evaluating AAMC PREview™ Scores with a Research-Only Performance Tool

May 2023

AAMC

Welcome, and thank you for joining this AAMC course series on **Local Validation Research: Evaluating the use of AAMC PREview™ Exam Scores at Your School – Is Local Validation Right for You?**

These courses were developed to provide information and guidance that may help you and your school decide whether and how to conduct your own research to evaluate the validity and usefulness of PREview scores in your admissions process.

In this course, **Using a Research Only Performance Tool (ROPT) to Evaluate PREview™ Exam Scores,** we will discuss the option of using a performance rating tool developed and administered specifically for research purposes. We will focus primarily on a rating tool AAMC developed as the medical student performance outcome in a PREview-related criterion-related validation study. You may, of course, also decide to tweak this tool or develop your own to best fit your research interests.

Learning Objectives

By the end of this Course, you will be able to:

1. Explain why your school may wish to study the relationship between PREview™ scores and medical student outcomes at your school.
2. Understand why your school may want to use a ROPT for validation at your school in comparison to other outcome measures.
3. Understand what a research only performance tool (ROPT) is.
4. Describe the steps needed to implement a ROPT study at your school.
5. Describe how to analyze results of a criterion-related validity study at your school.
6. Interpret these results and make inferences from these analyses.
7. Describe possible next steps to determine whether PREview scores add value to the admissions process.

AAMC

By the end of this course, our goal is for you to be able to:
1. Explain why your school may wish to study the relationship between PREview scores and medical student outcomes at your school.
2. Understand why your school may want to use a ROPT for validation at your school in comparison to other outcome measures.
3. Understand what a research only performance tool (ROPT) is.
4. Describe the steps needed to implement a ROPT study at your school.
5. Describe how to analyze results of a criterion-related validity study at your school.
6. Interpret these results and make inferences from these analyses.
7. Describe possible next steps to determine whether PREview scores add value to the admissions process.

Why Would You Want to Conduct Local Validity Research?

Stakeholders may find local research more compelling.

Different admissions data or performance outcome measures may be of higher interest than those used in AAMC research.

Local data can inform how PREview scores should fit into your school's unique mission and holistic review process.
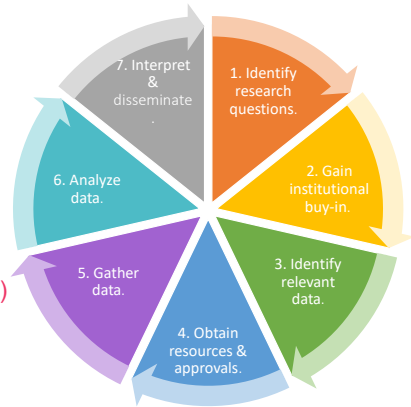
AAMC

Given that AAMC researchers have examined and documented results for research examining the relationship between the PREview score and some performance outcomes, why might you want to consider conducting local research at your school? There are several reasons.

- Stakeholders at your school may find local research more compelling in support of decisions regarding the use of PREview scores.
- Your school may have available or be able to collect-- and may value-- different performance and other outcome measures than those used in the AAMC study.
- Your school may have an interest in exploring how PREview scores should fit in to your school's mission and holistic review process. This may include determining the extent to which PREview scores help you better identify who will be successful at your school.

Study Checklist

- [ ] 1. Identify your research questions.
- [ ] 2. Gain institutional buy-in. (Revisited)
- [ ] 3. Identify relevant data.
  - [ ] Conceptual
  - [ ] Methodological (Study design)
- [ ] 4. Obtain resources & approvals. (Revisited)
- [ ] 5. Gather your data.
- [ ] 6. Analyze data.
- [ ] 7. Interpret & disseminate results.

AAMC

These are the high-level steps involved in studying how PREview scores may relate to your admissions process.
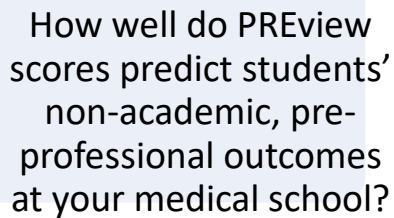
Course 101 addressed the importance of gaining institutional buy-in to the success of your research, so we won't discuss that further in this course. But, of course, to get leadership buy-in, you are going to have communicate the benefits of doing the research – that is, what research questions you are looking to address and why.

In Course 101, we also discussed the need to obtain resources including a project team that includes a data analyst, data analysis software, IT support to pull data, and IRB approval. We may touch on some of these topics in this discussion but won't go into them in depth.
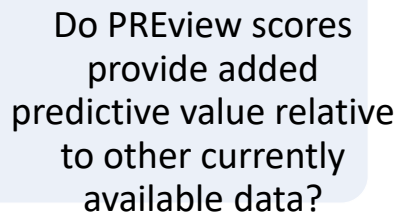
We will focus our discussion here on methodological and conceptual factors to consider regarding using a for research only performance rating tool as well as tips for gathering the data, analyzing, and interpreting the results.

Identify Your Research Questions

How well do PREview scores predict students' non-academic, pre-professional outcomes at your medical school?

Do PREview scores provide added predictive value relative to other currently available data?

© Association of American Medical Colleges

AAMC

The first step in any project is identifying research questions. These are the common RQs that we've heard from admissions officers.

How well do PREview scores predict students' non-academic, pre-professional outcomes at your medical school?

Do PREview scores provide added predictive value relative to other currently available data?

Identify Appropriate Outcome Data:
Existing Data Limitations

- Importance
- Conceptual relevance:
  o Measures of pre-professional competencies
- Methodological and practical concerns

AAMC

In Course 103, we discussed 4 things you should keep mind to identify appropriate outcome measures to answer your research questions: importance, conceptual relevance, and practical and methodological considerations. These all may be reasons that existing, available outcome data at your school are not appropriate, or at least sufficient, and why you may wish to use a Research Only Performance Tool.

It may seem obvious, but just because an outcome measure is available doesn't mean it is important. Important measures reflect critical aspects of student performance (e.g. grades, leadership, involvement). These measures should reflect what students accomplish, how they accomplish what they do (e.g., resilience, responsibility, ethics), as well as other important outcomes such as taking leave or dropping out.

Regarding conceptual relevance, recall that the PREview exam is intended to measure pre-professional competencies such as Service Orientation, Social Skills, Teamwork, Resilience and Adaptability, and Reliability and Dependability. It intentionally was designed not to measure some other things, especially academic and cognitive constructs that you would expect from tests of knowledge.

The outcomes that are available to you may not measure concepts or competencies

that are relevant to what the PREview exam was designed to predict.

- Traditional course outcomes (e.g., final grades) are mostly knowledge based and may have limited overlap with PREview-assessed competencies, especially to the extent they reflect primarily academic knowledge performance.
- Although your school may have some measures of important student pre-professional outcomes, the AAMC designed research only performance tool is a standardized validated measure of these competencies, and we will discuss more about this tool and its possible use at your school in the rest of this course.

Practical and methodological considerations may also be reasons for collecting your own performance data. To be useful, existing measures:

- Must be accessible – you have to be able to obtain the data.
- Must be in a useable form (e.g., numeric) or a form that you can readily make useable.
- Need to be available for all or most students, and it needs to reflect their individual performance. Group project scores or cohort levels statistics will be of no use.
- Must have variability in student scores.
- Must be reliably measured; they need to reflect a reasonably stable picture of student performance not something that would vary wildly from time to time or different evaluators.

So, let's discuss the AAMC's Research Only Performance Tool as an alternative outcome measure of medical student performance.

What is a Research Only Performance Tool (ROPT)?

| | Standardized evaluation of student performance/outcomes. | *Targeted raters evaluate students on relevant competencies.*
Uses a standard rating scale.
Given under similar conditions. |

| | Ratings are used for research only. | *Ratings DO NOT contribute to grades.*
Ratings are NOT shared for feedback.
More likely to reflect honest, accurate feedback.
More likely to observe variance in ratings. |

Because these data will be collected for research purposes, you will be more likely to get accurate, honest ratings and neither raters nor students need to worry about negative consequences, making ROPT one of the best performance outcomes to use.

AAMC

If existing data of meaningful student outcomes aren't available to you, or if they aren't conceptually relevant to what the PREview score measures or there are concerns about the availability or quality of that data, what do you do? The answer very well may be that you need to collect your own data, using a tool developed specifically for the purpose of answering your research questions about the PREview exam's usefulness for your school. This is what we mean by a Research Only Performance Tool.

- You define which students are rated (of course these should be students for whom you have PREview scores), and who rates them. Raters might include instructors, advisors, or clinical supervisors, but they could also include project team members or other peers. The type of rater you use should be consistent for all students.
- The same tool is used to evaluate all students that you include in your study. All students would be evaluated on the same items, using the same rating scales.
- The items should address important aspects of student performance that are relevant to the competencies the PREview exam was designed to measure.
- Ideally these evaluations are collected at about the same time in the student's journey, potentially after 1 or 2 years of medical school.

Because these data will be collected for research purposes, you will be more likely to get accurate, honest ratings because raters don't have to worry about negative consequences for either the student, because they won't affect grades or other aspects of the student's standing at your school, or for the rater, because the feedback provided does not need be shared with students.

Identify Appropriate Outcome Data: ROPT

- Importance ✓
- Conceptual relevance ✓
- Methodological and practical concerns

AAMC

So how might using a research only performance tool stack up against these considerations for appropriate outcome data? Pretty well actually.

Because you decide what to include in the measure, you can ensure that the questions asked reflect important and conceptually relevant outcomes.

Methodological and practical concerns come into play with respect to your study design , but first let's take a look at the AAMC designed tool we just mentioned. While you can certainly develop your own research instrument from scratch, you are free to use the content of this AAMC tool in your research, either as is, or as a starting point.

AAMC Developed an ROPT for M1-M2 Students that You Can Use!

- **Content:** Relevant professional competencies
- **Raters:** Faculty, preceptors, mentors or others who observe learner performance
- **Rating scale:** 5-point behaviorally anchored scales
- **Delivery:** Online survey

Research-Only Performance Ratings – Exercise

Competency Area 1

Reliability and Dependability

Please review the competency definition and the proficiency level descriptions and examples prior to providing your ratings.

Reliability and Dependability: Consistently fulfills obligations in a timely and satisfactory manner; takes responsibility for personal actions and performance.

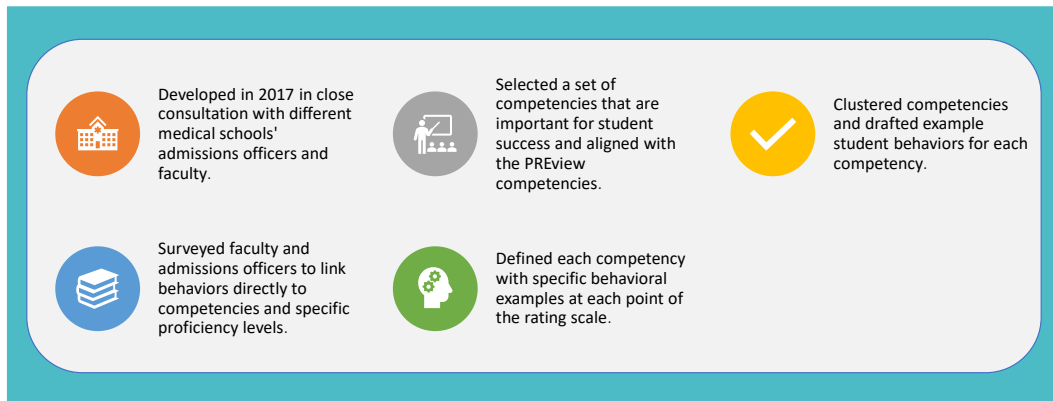| Proficiency Levels	Proficiency Level Descriptions and Examples
N/A	*Insufficient Information to Rate*
5 High	Completes assignments on time and comprehensively by holding self and others accountable; Prioritizes tasks effectively and thoroughly resolves conflicting responsibilities Examples: • Appropriately manages and attends to commitments, even when they compete with one another • Double-checks work and addresses potential errors with others
4 Medium-High	Completes assignments on time and comprehensively; Prioritizes tasks effectively and seeks alternative solutions to conflicting responsibilities Examples: • Contributes thoughtfully and comprehensively to discussions • Prepares and presents well-organized, accurate, and concise reports and cases
3 Medium	Completes assignments on time and satisfactorily; Prioritizes tasks effectively and satisfactorily addresses conflicting responsibilities Examples: • Meets deadlines for assignments • Arrives to class prepared to contribute to discussions
2 Medium-Low	Generally completes assignments on time and generally satisfactorily; May require assistance prioritizing tasks and managing conflicting responsibilities Example: • Communicates with faculty in case of late assignments or absences
1 Low	Occasionally fails to complete assignments on time and satisfactorily and sometimes misses an assignment; May fail to identify key priorities and/or address conflicting responsibilities Example: • May arrive to class without evidence of preparation

AAMC

The tool was developed by the AAMC specifically for research it was conducting on the criterion related validity of the PREview exam in order to explore how effectively PREview scores predicted performance on the 8 competencies it was designed to measure. Again, these included such things as Service Orientation, Social Skills, Teamwork, Resilience and Adaptability, and Reliability and Dependability. This means the content relevancy of the tool is high, and this tool is likely a more direct measure of PREview exam competencies than any other rating scale.

The PREview exam was designed to measure these competencies because they represented important aspects of medical student performance that were not being as directly addressed by more academically focused admissions data such as the MCAT® exam or UGPA.

The ROPT is linked and available to download right on the site page where you have accessed these modules.

How was the AAMC ROPT Developed?

Developed in 2017 in close consultation with different medical schools' admissions officers and faculty.

Selected a set of competencies that are important for student success and aligned with the PREview competencies.

Clustered competencies and drafted example student behaviors for each competency.

Surveyed faculty and admissions officers to link behaviors directly to competencies and specific proficiency levels.

Defined each competency with specific behavioral examples at each point of the rating scale.

This research tool was developed by AAMC experts in test development and validation, using a strong, systematic methodological approach

- It was developed in close consultation with subject matter experts from different medical schools.
- As mentioned, the competencies included were selected because they were identified as important for student success and aligned with the PREview competencies.
- After the competencies were defined, example student behaviors for each were developed.
- These were reviewed by subject matter experts who ensured that behaviors were linked directly to competencies and specific proficiency levels.
- As a result, behavioral examples were defined for each point of each rating scale.

Buy-In, Resources, and Approvals

- **Leadership support and rater buy-in is critical when using an ROPT.**
  - Active data collection will require more resources.
- **May need to investigate validity of resources.**
  - Project team
  - Data analysis software
  - IT or other support
  - Survey software or administrative help to hand enter data
- **Might require Institutional Review Board approval.**

So, if you decide to use the AAMC tool, you know you have a well-designed, relevant measure of pre-professional competencies. If you decide to create your own (which is beyond the scope of what we can cover here today) you at least have a starting point and a model for what such a tool might look like and how it might be developed.
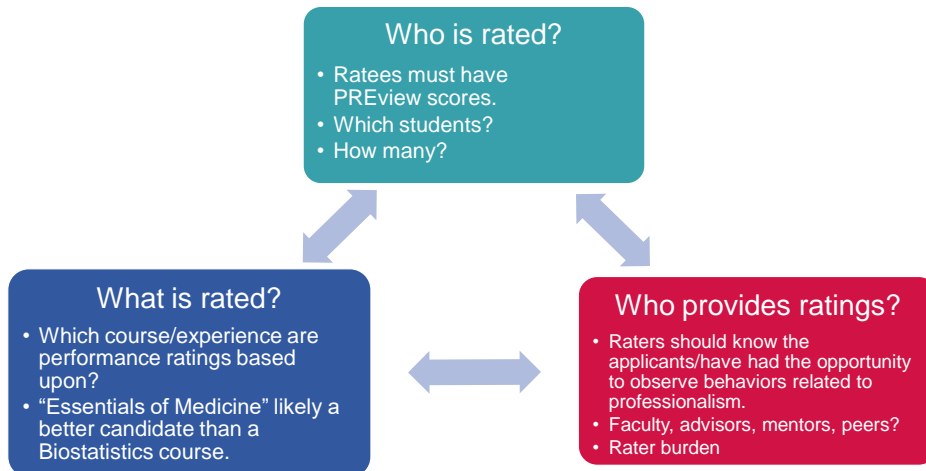
Before we get into study design, we want to remind you of the importance of getting leadership buy-in, as well as needed resources and approvals

- Leadership support and buy-in will be particularly critical if you decide to use a Research Only Performance Tool. Gathering your own performance data will require more time and resources than using what already exists, including your time, the time of other staff to collect and analyze data, and of course the time of raters who are asked to provide the data. It will be important to have one or more influential individuals at your medical school champion the value of this research – and communicate its importance to any individuals you may need to leverage to complete the research. For example, if you collect performance

ratings from faculty, they may be reluctant to take on the additional burden of rating students. Having an internal champion can help combat this potential roadblock.

- Gathering the right human capital and resources also will be important for the success of this research:
  - We mentioned in Course 101 that your project team might need to include a data analyst with knowledge/experience in data collection, statistical analyses, and interpretation. We also mentioned the possible need for IT help to access data and for software (e.g., Excel, SPSS, SAS, R) to analyze the data
  - In this instance, where you are collecting your own data, you also need a way to collect it. These days, this is often accomplished through some form of survey software, but you could decide to send out paper forms and then hand enter the returned data. Either way, you may need to find extra resources to accomplish this.

- Finally, it is important to know from the onset of this work that you will very likely need to get your research approved through your school's Institutional Review Board or IRB, because you are collecting information about student from students or employees.

Study Design – The Big Questions

Who is rated?
- Ratees must have PREview scores.
- Which students?
- How many?

What is rated?
- Which course/experience are performance ratings based upon?
- "Essentials of Medicine" likely a better candidate than a Biostatistics course.

Who provides ratings?
- Raters should know the applicants/have had the opportunity to observe behaviors related to professionalism.
- Faculty, advisors, mentors, peers?
- Rater burden

© Association of American Medical Colleges

AAMC

The three questions here are intricately connected. Who is rated may seem obvious-- you want ratings on students who have preview scores. But which students?

An important factor in this decision is to whom do you want to generalize your survey findings with respect to the decisions you make.
- If you want to use your results to make changes to how you admit next year's class of students, it may make sense to limit your study to only your most recently matriculated students.
- But if only certain classes provide opportunities for students to demonstrate the competencies targeted by PREview, then only students who have taken those classes would be rated. If these classes are taken by all students, great. But if they are electives taken only by some students, the results you find may not be as broadly generalizable to other students or applicants
- Another determining factor is on how many students will need ratings, that is, the sample size for the study. This gets at the issue of statistical power, which we will discuss on the next slide, but decisions you make on who and what to study may limit your ability to include enough students at a given time, requiring you to spread data collection across more than one year.

The second consideration of course is who will provide the ratings?

- The easy part of this answer is that **raters should know the applicant well and have had the opportunity to observe behaviors that fall under the competency categories of the ROPT (measured by PREview).**
- But should those raters be faculty that teach them, advisors or mentors, or even peers?
    - Peers may be the people who know students best with regard to the competencies addressed by the PREview assessment, but using peers can add a lot of complexity to the research with regard to communication and possibly raising expectations that results should be shared with those who are evaluated.
    - Faculty should be familiar with the performance of students specific to the classes they teach, but may not be familiar with their performance overall. Faculty teaching different classes may see different aspects of students, and certain classes may be more likely to elicit PREview exam-relevant behaviors than others.
    - Advisors, coaches, or mentors work with students more broadly, but may or may not know specifics about the strengths or weaknesses students display with respect to individual classes or the competencies you are looking to evaluate.
    - Another consideration is the burden placed on raters, that is, how many students they are asked to rate. It would be unreasonable to expect a teacher with a lecture hall of 50 students to know them all, much less ask them to rate all 50. At the same time, limiting the number of students a person is asked to rate to 10, let's say, may limit the number of students who can be rated, reducing your study's sample size and power.
    - **Ultimately you will have to use your knowledge of your school to determine what is best in terms of both what will result in the best ratings and what is practical and will be acceptable.**

The third consideration is from what course or on what experience you will gather ratings. As we have mentioned in many of these courses, the outcomes you use should be conceptually related to the competencies the PREview test is designed to measure. It follows then that the courses or experiences for which you gather ratings should be those that are likely to elicit behaviors that reflect those competencies.

- A Biostatistics class in which students work independently and are graded primarily on test performance may be a poor choice for gathering this feedback, because faculty who teach those courses may not be as familiar with their students professionalism.

- On the other hand, a course addressing societal concerns and in which students

are required to interact and work together may be a perfect setting for observing performance on conceptually relevant competencies.

Study Design – Sample Size

- Large sample sizes (Ns) may not be available, practical
- Small sample sizes limit ability to find relationships that exist because they lack statistical power

MOCK DATA: FOR ILLUSTRATIVE PURPOSES ONLY

Measure	AAMC Observed r	Power (N=85)	Needed #, Power=0.70
AAMC Survey Performance Measures (M2 sample)			
Resilience & Adaptability	0.12	0.20	427
Teamwork	0.16	0.31	239
Ethical Responsibility	0.19	0.42	169
Reliability & Dependability; Capacity for Improvement	0.21	0.49	138
Social Skill; Service Orientation	0.33	0.88	55
Cultural Competence	0.36	0.93	46

© Association of American Medical Colleges

AAMC

Sample size is an important consideration in your study design, that is, how many records or cases of data do you need (how many students with PREview data do you need data for)?

Practical concerns such as the total number of students you have in a given cohort, or the number of students you can reasonably expect a rater to evaluate can limit your sample size. However, small samples represent a risk to validity research because they limit your ability to find statistically significant results that exist in reality unless the true underlying relationship between your two variables of interest is very strong. This has to do with a consideration called *statistical power*.

Stated simply, power is the likelihood of correctly finding a relationship that exists. Power is affected by the strength of the relationship and the sample size in your study. The flip side of power is called "Type II Error;" that is, erroneously concluding there is no relationship when in fact a relationship exists. So, with small sample sizes, you will have less power and are more likely to conclude that no relationship exists between PREview scores and other variables, even when it does.

Using outcome correlations (r) from earlier AAMC research, **this table** shows the

power that would result from a moderate sized study (N=85) and the study sample size you would need to have a 70% chance of finding a statistically significant relationship between PREview scores and that variable (that is, power=0.70). If you are interested in exploring any of these relationships in your own study, this table can provide some guidance regarding the number of study participants (with data on both PREview scores and the other measure of interest) to include. Depending on the relationships of interest to you and the size of your matriculant class, it may take several cohorts/years of data collection before you can have enough data to establish a reliable correlation.

Study Design – Other Topics

- How to collect?
  - Medium
  - Type of survey
- Linking data
- Timing
  - Avoid busy times.
  - Avoid timing such that the act of rating someone using the ROPT could affect their course grades and vice versa.

MOCK DATA: FOR ILLUSTRATIVE PURPOSES ONLY

AAMC ID	Stu_ID	Name	Race
******0905	95-0602	Willima Smith	White
******8424	98-3211	Billy Smith	Black
******3373	99-1299	Will Smith	Black
******2468	88-2113	G. William Smith	White

AAMC ID	Date	Name	PREview score
******0905	9/22/2022	William H. Smith	5
******8424	10/7/2022	W. Elizabeth Smith	7
******3373	11/3/2022	Everett W. Smith	8
******2468	5/17/2022	George Smith	4
******2468	12/12/2022	George W. Smith	6

Stu_ID	Name	Interview - Teamwork
95-0602	Smith, Bill	1
98-3211	Smith, Wilhelmina	3
99-1299	Willis Smith	2

Ultimately, all data for each participant will need to be linkable.
- One row per participant.

AAMC

As we touched on briefly earlier, another consideration is the medium you use to collect your data. For example, you might choose to use an internet-based survey software (such as SurveyMonkey or Qualtrics), email Excel forms for people to complete, or send out paper forms.

- Survey software will generally be your best bet, but it needs to be something you have access to (either directly or through others) and know how to use. In addition, the people you survey will need to have email and internet access – something that probably won't be a problem at most medical schools. However, you will also need to explore whether your IT department will allow respondents to receive the survey emails and access the survey site

- Paper or Excel surveys are really only practical when you have a small number of raters, because you will need to physically transcribe the data (in the case of paper) or combine multiple files (in the case of Excel). In addition, these two approaches are likely to result in lower response rates than an internet-based survey

Another consideration is whether you will seek input from a single rater or multiple raters per person (for example 3 professors).

- The first approach is methodologically easier and easier for some survey

software to handle. On the downside, you are getting one person's perspective, and the ratings they provide may reflect as much on the person doing the rating, or the class or experience in which they interacted with the student, as the person being rated. Put another way, the ratings may not be as reliable. Also, if that one rater does not respond, you are missing rating data.

- Gathering input from multiple raters is more complex, and can make it more difficult to avoid overburdening some raters. On the other hand, you can (and should) aggregate the multiple ratings of a person into averages for each competency, which should result in more reliable performance scores. In addition, if you seek input from multiple raters per student, you will have more useable data even if a particular rater fails to respond.

Linking is something we have addressed in other courses where we discussed using existing admissions or performance data. Simply put, if you can't link the performance data you collect to PREview scores (and possibly other) data, you may have wasted your time and that of others.   Some shared variable or variables unique to each candidate but shared across sources of data is critical to your research success.

- Names may be used but are often problematic Multiple people may have the same name, the same person may have different variations of their name in different data sets, or there may be typos in one version of a name, as shown in this example where every person might go by Willy Smith.
- AAMC ID is common and should be available for all candidates
- It also may be necessary to have more than one matching variable and to match different data sets on different variables. As shown here, the interview data (green) had Student ID, while the PREview data did not (Orange; people taking PREview assessments were not students yet) but did have student ID. Fortunately, the demographic data from student files had both variables, otherwise it may have been difficult to accurately match the test and performance data.
- Generally speaking it is not usually difficult to link internet survey data to specific respondents, as long as you plan it out ahead of time.

Finally timing of your data collection is something to consider. Obviously, you won't want to ask for ratings from people when they are not likely to be available, such as over summer or during spring breaks. You also should seek to avoid times during the year when raters are most likely to be busy, such as while students are taking final exams or professors are assigning final grades. This is a good time of year to avoid also so that the act of grading someone does not affect the ratings provided and vice versa.

Gathering Your Data

1. Finalize the list of students at your school with PREview scores for whom you want to collect performance data.
2. Establish buy-in to recruit faculty/coaches/mentors/peers to participate and complete ROPT ratings.
3. Determine the list of raters and the students they are assigned to rate.
4. Develop the ROPT survey.
5. Administer the ROPT survey to at the agreed upon date.
6. Have faculty rate students.
7. Collect ROPT survey ratings and analyze results.

AAMC

At the risk of oversimplifying, this slide describes the basic process of gathering your data, once you have decided who you want to survey, about whom, and about what. To gather your data you should:

1. Finalize the list of students at your school with PREview scores for whom you want to collect performance data.
2. Establish buy-in to recruit faculty/coaches/mentors/peers to participate and complete ROPT ratings.
3. Determine the list of raters and the students they are assigned to rate.
4. Develop the ROPT survey.
5. Administer the ROPT survey to at the agreed upon date.
6. Have faculty rate students.
7. Collect ROPT survey ratings and analyze results.

Analyze Data

- Data merging
- Data cleaning
- Identify appropriate analysis

Research Question	Analysis
How well do PREview scores predict students' non-academic, pre-professional outcomes at your medical school?	• Correlation, regression • Risk ratio
Do PREview scores provide added predictive value relative to other currently available data? (Incremental Validity)	• Multiple regression • Logistic regression

This is where your data analyst or analyst team really earn their money. There is only so far we can go in advising you what to do when we don't know the data you will have, but we can provide an overview of some important considerations

Data-merging: putting all of your data, from different sources, together in a single line of data based on one or more linking variables
- o handling one-to-many matches – which do you use?
    - o Multiple (disagreeing) demographics.
    - o Multiple occurrences of PREview scores.
- o Handling missing data/unmatched cases.

For data-cleaning, you need a plan for:
- o Missing or out of range (impossible) values.
- o Reasons for dropping applicants from the analysis – e.g., applicants who spent less than 1 minute on the rating exercise, or provided "flat line" ratings.

- How well do PREview scores predict students' non-academic, pre-professional outcomes at your medical school? You will be looking at the relationship between PREview scores and your outcome measures
    - Generally, this analysis would look at:

- - Correlation coefficients to understand the strength of the relationships.
    - Regression lines (equations) to understand the nature of the relationships.
    - If the nature an outcome is dichotomous (binary), such as dropped out/did not drop out, you may also wish to look at risk ratios.
- Do PREview scores provide added predictive value – that is, incremental validity, above and beyond other admissions data?
    - This analysis of course would require you to have not only PREview and outcome data, but also all other admissions data for the students in your analysis. *(e.g., UGPA, MCAT® scores, interview ratings, and MMI ratings)*
    - The analysis you conduct would be multiple regression, or multiple logistic regression
    - The primary question you would examine is whether PREview scores helps predict an outcome after all other admissions data have been considered.
    - A secondary question could be whether PREview scores can be used to replace some other data you currently use for admissions
    - To answer either of these questions, we would urge you to run these analyses on multiple relevant outcomes.

Interpret Results - Correlation

Research Question	Analysis
How well do PREview scores predict students' non-academic, pre-professional outcomes at your medical school?	• Correlation, regression • Risk ratio

MOCK DATA: FOR ILLUSTRATIVE PURPOSES ONLY

	PREview scores	Adaptability	Teamwork	Resilience	Social Skills
PREview scores	1.00				
Adaptability	0.14	1.00			
Teamwork (peer)	0.23	0.19	1.00		
Resilience (peer)	0.21	0.41	0.18	1.00	
Social Skills	0.22	0.19	0.47	0.21	1.00

AAMC

An important consideration about a comparison of differences or a correlation coefficient is whether it is "statistically significant." In your research, very few of the differences or correlations you find will exactly equal 0. Statistical significance pertains to the question of whether a given result is far enough away from zero to represent something that is a real, potentially meaningful finding. An in-depth discussion of significance testing is beyond the scope of this session, but the question it answers is "what is the likelihood that this result is due to chance?" When that likelihood is less than a certain amount (commonly 0.05, or 5%), most researchers are willing to accept that the result is different from 0.

If you find statistical significance, you can then look at the size and direction of differences with respect to the research question about changing applicant pools, or the size and pattern of the correlations for the question about PREview and other applicant data.

In order to provide a complete story and round out the checklist steps we shared earlier, we will provide examples of the results you might find for the two research questions we discussed.

The results of the analyses for the first research question posited earlier might look something like this. (Note, these results are made up for illustrative purposes).

PREview scores, which is in the first row and column of the table, have lower correlations with ratings of Adaptability, and moderate correlations with Social Skills and peer ratings of Teamwork and Resilience.

Interpret Results - Regression

Research Question	Analysis
Do PREview scores provide added predictive value relative to other currently available data? (Incremental Validity)	• Multiple regression • Multiple logistic regression

MOCK DATA: FOR ILLUSTRATIVE PURPOSES ONLY

Dependent Variable: Faculty Eval - Dependability					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-2203.78	2950.242		-0.75	0.46
Personal Statement	.0002	0	0.06	0.74	0.46
Interview	0.30	0.115	0.23	2.64	<.001
GPA	.0001	0	0.23	2.64	<.001
2 (Constant)	-959.45	2764.635		-0.35	0.73
Personal Statement	.0003	0	0.03	0.34	0.73
Interview	0.05	0.121	0.04	0.44	0.66
GPA	.0001	0	0.18	2.31	0.023
PREview	0.47	0.107	0.40	4.39	<.001

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.33	.11	.09	1.02	0.11	4.97	3	121	0.003
2	.48	.23	.21	0.95	0.12	19.26	1	120	<.001

AAMC

In the case of the second research question (again results are made up for illustrative purposes only) interview and GPA do a good job of predicting Dependability, without PREview in the mix.

After adding PREview scores, not only does they add to the prediction of Dependability (more than doubling the R square or variance accounted for) it is the best predictor, and in fact, you do a good job of predicting Dependability without Personal Statement or Interview data.

If you saw a pattern like this across multiple important outcomes, you could build a case for dropping one or both of those resource intensive types of applicant data.

In Summary:

- Including important, conceptually related, and methodologically sound performance outcomes will be key to the success of this research

- While using an ROPT can be your best bet for achieving those goals, it will also require more work than using existing/available outcomes. For this reason, gaining institutional support and buy-in is particularly important here. Such support is needed if you are going to place demands on other people's time.

- Potential analysis techniques will depend on the research questions you want to answer. The ones we discussed in this course are most likely to rely on Correlations, Multiple Regression, and Multiple Logistic Regression.
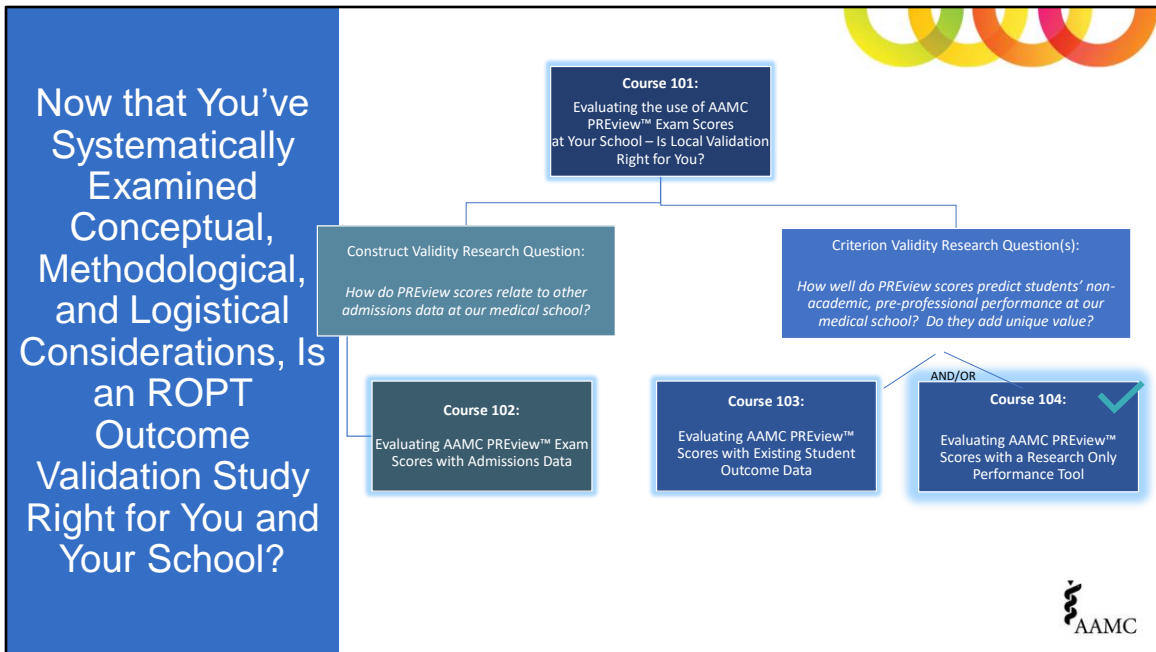
Now that you've systematically examined conceptual, methodological, and logistical considerations, is an ROPT outcome validation study right for you and your school?

If you are skeptical that you will have the resources to do this kind of research, and you haven't already, check out Course 103, which is about using existing course outcomes to do an outcome validation study.