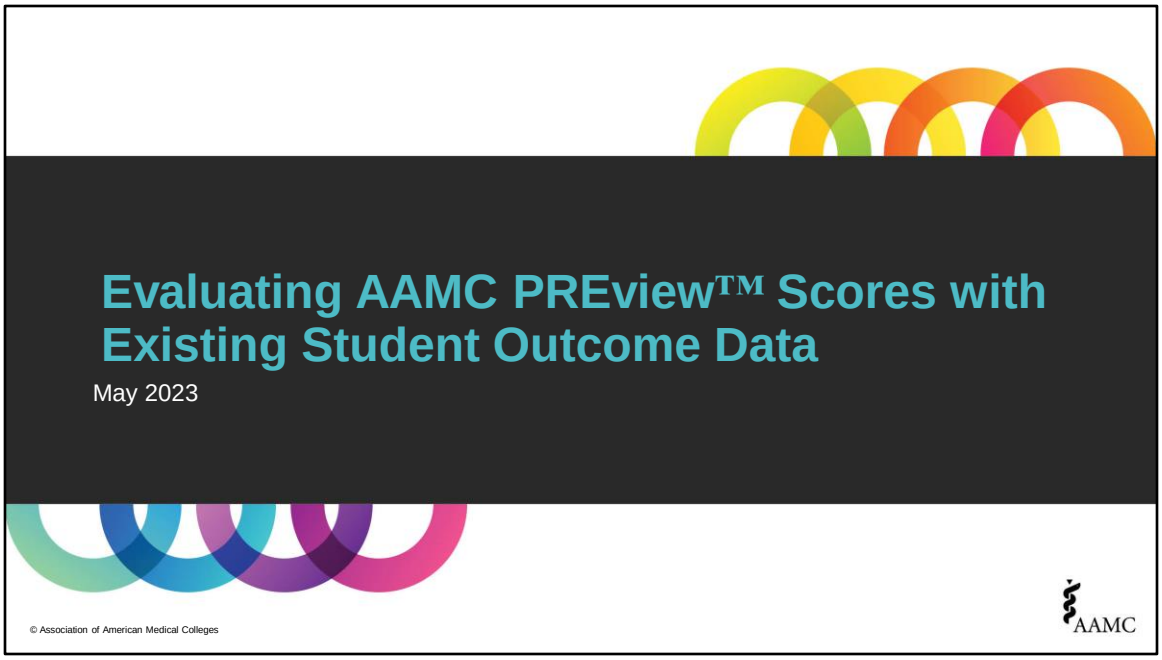


Welcome, and thank you for joining this AAMC course series on **Local Validation Research: Evaluating the use of AAMC PREview™ Exam Scores at Your School – Is Local Validation Right for You?**

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, we will discuss **Using Existing Outcomes to Evaluate the AAMC's PREview™ Exam at Your School via a Criterion Validation Study,** we will go over important information about considerations for using existing outcomes, like course-based grades, in a criterion validation study for evaluating the PREview exam at your school.

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. Review and identify outcomes that are appropriate to include in a validity study.
3. Describe how to perform analysis for a criterion-related validity study at your school.
4. Interpret these results and make inferences from these analyses.
5. Describe possible next steps to determine extent to which PREview scores add value to the admissions process.

AAMC

By the end of this course, our goal is for you to be able to:

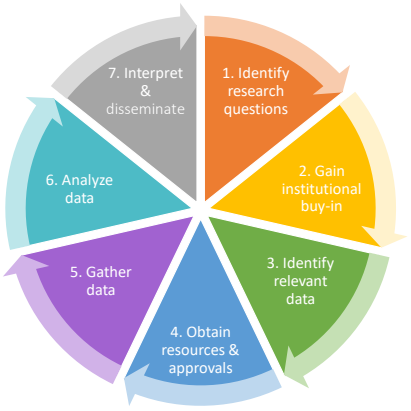
1. Explain why your school may wish to study the relationship between PREview scores and medical student outcomes at your school.
2. Review and identify outcomes that are appropriate to include in a validity study.
3. Describe how to perform analysis for a criterion-related validity study at your school.
4. Interpret these results and make inferences from these analyses.
5. Describe possible next steps to determine extent to which PREview scores add value to the admissions process.

Given that AAMC researchers have examined and documented results for research examining the relationship between the PREview exam 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 value, have available, or be able to collect different performance and 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.*
- ❑ 3. Identify relevant admissions data.
  - ❑ Methodological
  - ❑ Conceptual
- ❑ *4. Obtain resources and approvals.*
- ❑ 5. Gather your data.
- ❑ 6. Analyze data.
- ❑ 7. Interpret & disseminate results.

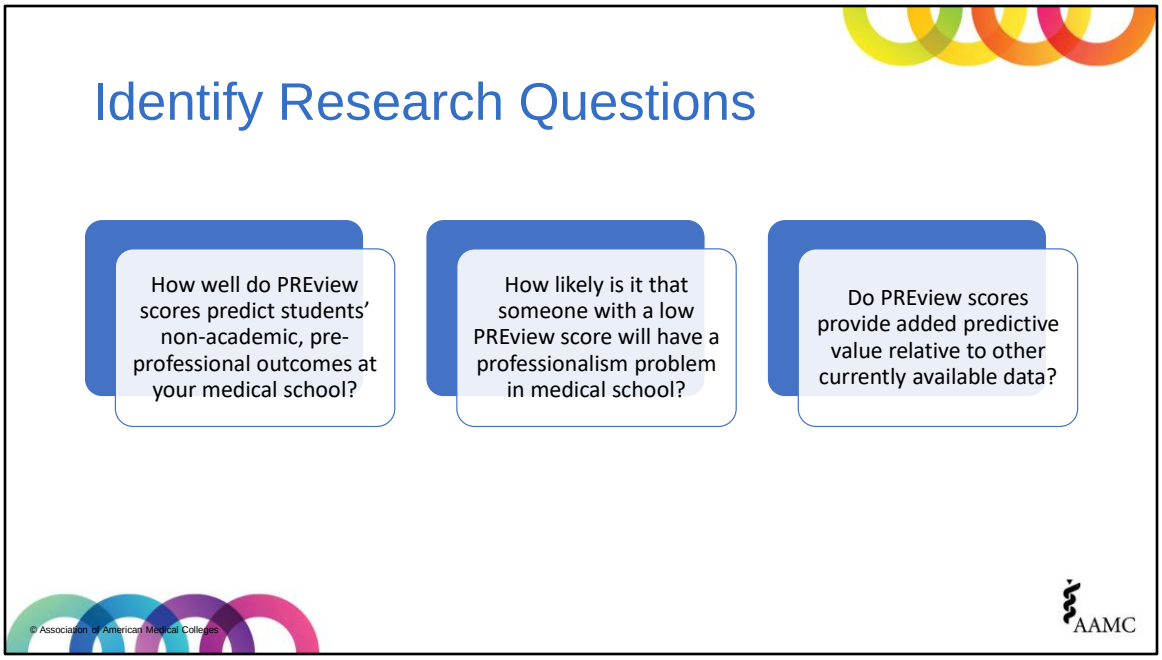


AAMC

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/leadership 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 to communicate the benefits of doing the research; that is, clarify 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 but won't go into them in depth.



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?
- How likely is it that someone with a low PREview score will have a professionalism problem in medical school?
- Do PREview scores provide added predictive value relative to other currently available data?

Identify Appropriate Outcome Data: Considerations

- Importance
- Conceptual relevance
  - Measures of pre-professional competencies
  - Measures reflecting multiple or unknown competencies
- Methodological and practical considerations

AAMC

To identify appropriate outcome measures, there are 4 things you should keep in mind: importance, conceptual relevance, methodological considerations, and practical considerations.

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 <u>what</u> students accomplish, <u>how</u> they accomplish what they do (e.g., resilience, responsibility, ethics), as well as other important outcomes, such as taking leave or dropping out.

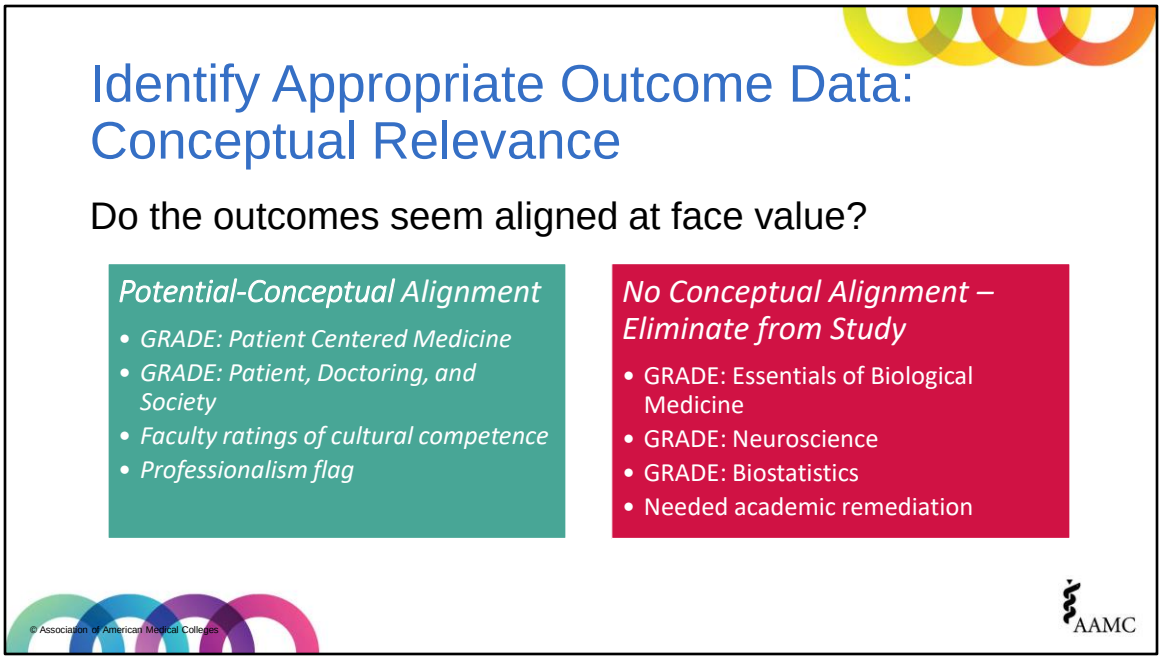
Regarding conceptual relevance, recall that the PREview exam is intended to measure pre-professional competencies such as Service Orientation, Social Skills, Teamwork, Resilience and Adaptability, and Reliability and Dependability. It intentionally was designed not to measure some other things, especially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

- If the data are readily available, it may be *interesting* to know how PREview scores relate to academic/cognitive performance; however, examining the extent to which PREview scores predict something they are not intended to measure

probably should not be your primary focus.

- It makes more sense to see whether the PREview exam is doing its job by predicting non-academic aspects of performance, such as how well a student works with peers and the public (including those from different and diverse backgrounds), whether a student keeps their commitments, is ethical and follows rules, or tolerates stress and uncertainty.
- There also may be other outcomes where you are not sure what exactly is being reflected. For example, does a student dropping out reflect academic or professionalism struggles, or both. In cases like this it may, be informative to see how strongly PREview scores relate to these outcomes.

Of course, practical and methodological considerations are just as important. You can identify the best, most conceptually relevant construct, but if you can't get data that measures it well, it will be of little use. We will discuss these considerations in more depth shortly.



This slide takes a slightly deeper dive into the topic of conceptual relevance. It shows outcome data your school may have available that may be more and less useful from the perspective of conceptual relevance.

As discussed in Course 101, PREview scores reflect examinees understanding of effective and ineffective pre-professional behaviors. In other words, you'll be looking for performance outcomes that are conceptually relevant that measure these or similar competencies. Outcomes that primarily reflect cognitive performance, such as traditional test-based academic grades (particularly in classes addressing cognitively difficult topics), likely are not very conceptually relevant to what the PREview exam measures. Grades in other courses, or at least components of those grades, may incorporate measures of professionalism competencies that are more relevant to PREview competencies. And of course, your school may have measures of student performance other than grades, such as faculty or peer ratings of students, or records of professionalism concerns.

Identify Appropriate Outcome Data:
Methodological and Practical Considerations

Data must:

- Be accessible.
- Be useable:
  - In existing form.
  - Need recoding/aggregating/transforming.
- Be available for most students.
- Reflect individual performance.
- Have variability.
- Be reliable.

AAMC

First, to be useful, the data must be accessible; that is, it is either data you have, or data you can retrieve in a reasonable time and with a reasonable amount of effort. If you don't have relevant, accessible data, Course 104 addresses how to collect your own data using a measure specifically designed for your research purposes.

Next is the question of the whether the data are useable, or what it will take to make them useable. Grades, whether individual course grades or GPA, likely will be useable in their existing form. Other, more qualitative data, such as peer or professor or patient recognition, may need to be converted to numeric scores to be useable.

By available, we mean that most students should have it. Measures that only a small subset of students have likely will not prove very useful because the number of cases with both PREview scores and outcome data may be too small for you to find relationships that exist and students with those data may not be representative of your general student population.

The data you use must be available at the individual student level. Measures that reflect group performance, such as group project grades, may not reflect what role each individual played in the final result or how they performed, if they all received

the same, group score.

There also must be variability in the scores. If almost everybody has the same score on something, it will not be related to other measures.

- With a dichotomous or binary variable (one that can have only one of two values), this relates to the concept of base rate. For example, if only one or two students per year are discharged for ethics violations, there would not be enough instances of this discharge status (discharged or not) to constitute a useful measure. On the other hand, if a higher number of students are subjects of ethics reviews, discussions or warnings, whether students were subject to a review or received a warning may be measures with enough variability for analysis.

Finally, to be useful in this research, the predictors you use should be reliable or fairly stable measures over time.

- We acknowledge that it may be difficult to evaluate the reliability of the data that are available to you; at times the best you may be able to do is consider the reputation that data have with people who use it (e.g., "We have this field, but it isn't very useful," or "It isn't updated very regularly.").
- For qualitative data that may need to be transformed to a numeric form (e.g., peer or professor or patient recognition), it is important that the measure be reliably quantifiable. That is, not only can it be converted to a numeric score, but it can be done in a way where different people will agree on how things are quantified (come up with the same score).
- It's useful to examine course outcomes from this perspective as well. For courses outcomes, is there a rubric being used? Is it used consistently? For other types of data, are outcomes reported consistently and updated regularly? All of these factors influence the reliability of any outcome.

In summary, the extent to which PREview is related or unrelated to your outcome data will depend both on whether they measure similar competencies, and on other characteristics of the data such as reliability and variation.

Choosing Outcomes Based on Relevance and Methodological Appropriateness

Outcome Category	Description	Conceptually Relevant	Important	Accessible	Useable	Available	Individual	Variable	Reliable
Course Grade Based on Test Scores	Final grade in 1st or 2nd-year course(s) that reflect facts/knowledge learned	Low	High	High	High	High	High	High	High
Course Grades Incorporating Participation, Behaviors	Final grade in 1st or 2nd-year course(s) that incorporate patient and/or team interaction	Moderate -High	High	High	High	High	High	Moderate -High	Moderate -High
Faculty Evaluation (Subscores)	Subscores focusing on relevant competencies (e.g., professionalism)	High	High	High	Moderate	High	High	Moderate	Moderate
Professionalism Flag	Recorded incidents in which students had professionalism violations	High	High	Low-Moderate	Moderate	High	High	Low	High
Peer Feedback/Rating(s)	Students evaluate one another based on performance in small group discussion/activities	High	High	Low	Low	Low-High	High	Low	Moderate
Objective Structured Clinical Examination (OSCE) Rating(s)	Students practice and demonstrate clinical skills in a standardized medical scenario, are evaluated by faculty	Low	Moderate	High	High	High	High	High	Moderate -High
Professional or Student Leadership/Involvement	Students lead or take active role in professional or student organizations or activities	High	High	Low	Low	Low	High	High	Low
Turnover	Whether students drop out or take leaves of absence	Moderate	High	High	High	High	High	Low-Moderate	High

AAMC

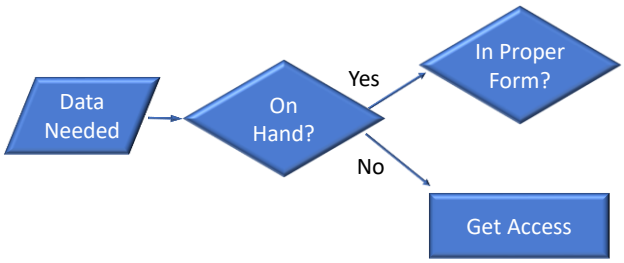
This is an example of an evaluation of some possible outcomes that may exist at your school. Of course, your school may value these differently, rate them differently, or have other available outcomes.

Let's look at a few rows.

- 1. Course grades or GPAs based primarily on test scores are likely easy to access, in a useable form, available on all students, variable, and measured reliably, but these outcomes are low in conceptual relevance in that they reflect cognitive ability and academic competencies that the PREview assessment was not deliberately designed to measure.
- 2. Grades in courses that incorporate measures of competencies like teamwork, social skills, or reliability are high on most characteristics, but may be only moderately relevant depending on how much tests of knowledge are incorporated in the final grade. The variability of grades will be affected by how non-academic performance is measured, and how seriously professors take the task of grading that performance.
- 3. Professionalism flags can be highly conceptually relevant, but may not be in an accessible format or even more importantly, reflect enough variability to analyze given the potentially small number of people who are flagged (<1% of a class).

- 4. Professional or Student Leadership involvement could be a highly relevant and important outcome to examine, but may face issues in terms of accessibility, usability, availability, and reliability as a metric for professionalism.

Gather Your Data – Accessible/Useable



- Proper Form
  - Data should be transformed into numeric scores.
  - The process of transforming should be systematic and reliable.
- Access
  - Get permission to obtain data.
  - Be specific about data needs.

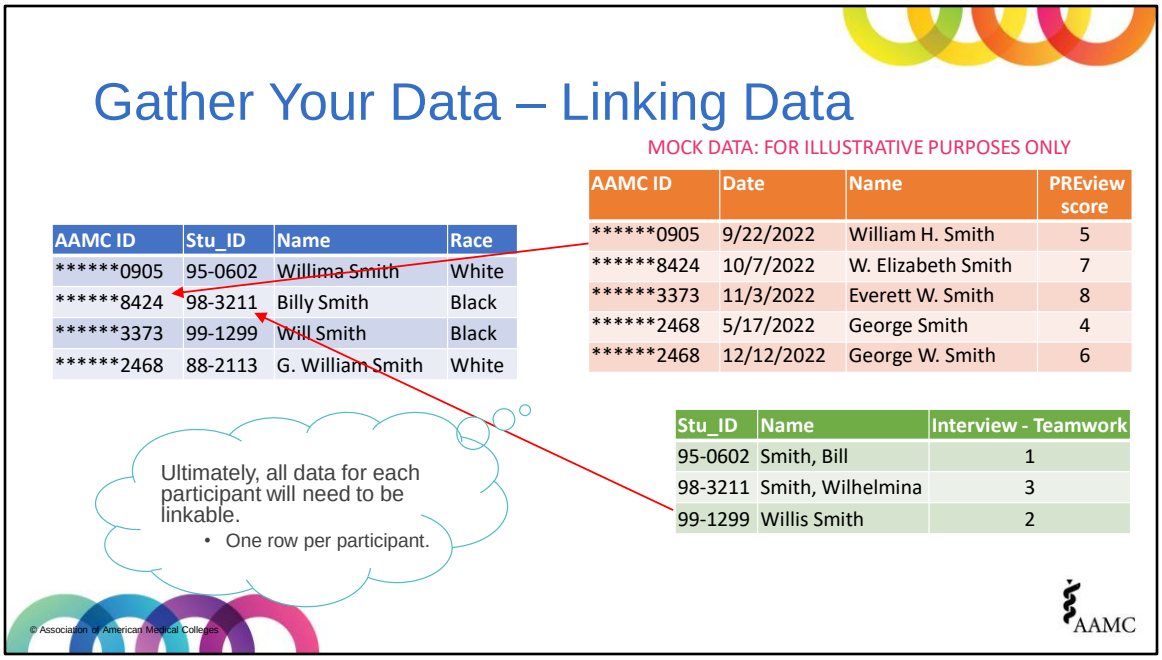
AAMC

We've discussed the issue of data being accessible and useable when we talked about Identifying Relevant Admissions Data.

- Some data you need may already be on hand, meaning that you can directly download it or otherwise access it.
  - If you can directly access the data, the next question is whether it is in the proper form. That is, is it a numeric variable where there is some order to it –for example higher scores which mean more or better of something.
  - This would be the case with measures such as UGPA, MCAT® scores, and PREview scores. It may be the case with interview scores or other measures as well.
    - However, there may be other admissions data where this isn't the case, for example if your school doesn't systematically assign scores to letters of recommendation, personal statements, or records of experience, or perhaps other data in student files. For these data to be useful in the analyses we are discussing, they need to be transformed into numeric scores and this needs to be done systematically and reliably. In other words, different people

scoring the same information should be looking for the same things and should wind up with similar scores. If different raters differ too much, the scores may be more reflective of the raters than the applicants.

- Other data you may need to get access to from other sources, perhaps your IT department, AAMC, or another assessment vendor. Once you get the data, the process is like what we described above. However, there are two issues in this case that are important to consider
    - First, you may need permission to get data from other sources, so that needs to be part of your planning and timeline
    - Second, you must be very specific about what data you need. This includes
        - Who – for which applicants, which time period.
        - In what format? Do you want the data in Excel, Access, Word, or some other file format? How do you want duplicate cases handled (e.g., use most recent record)?
        - Which variables at what level or in what format– for example do you need item or dimension level data or just a total score? Will the data be in a numeric format or not? Will you need a key or data dictionary to understand what different data mean?
            - An important point to consider here (and with the data you have on hand too) is how you are going to link together different pieces of information about an applicant. We are going to cover this topic more in the next slide, but the important point when requesting data from others is that you will have to figure out what linking variables are available and be sure to request ones that you need.



Another important consideration for gathering your data is being able to link it.

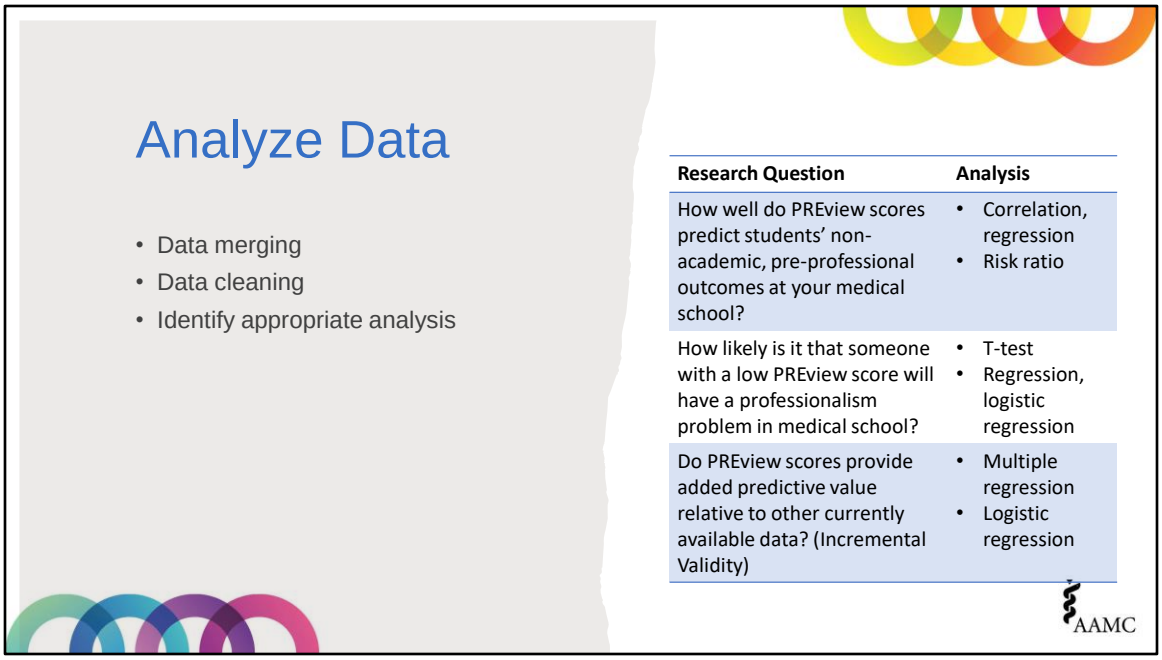
All this data may be worthless if you are not able to connect data from different sources. Some shared variable or variables unique to each candidate but shared across sources of data is critical to your research success.

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.

AAMC ID should be common and available for all candidates, and other forms of ID include AMCAS ID or email address.

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 were not students yet) but did have AAMC 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.

You or your data analysts will likely have dig in deeply to understand what is available and what problems might occur.



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.

- Handling one-to-many matches – which do you use?
  - Multiple (disagreeing) demographics,
  - Multiple occurrences of PREview scores,
  - Multiple occurrences of same outcome variable – aggregating or selecting one?
- Handling missing data/unmatched cases.

For data-cleaning, you need a plan for:

- Missing or out of range (impossible) values,
- Reasons for dropping applicants from the analysis.

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 in outcome is dichotomous (or binary), such as dropped out/did not drop out, you may also wish to look at risk ratios.

How likely is it that someone with a low PREview score will have a professionalism problem in medical school? Identifying a "low" PREview score will depend on your own school's unique considerations in terms of your admissions and selection goals.

- In its simplest form, you can answer this question by simply defining what you mean by a low PREview score and the computing the average professionalism scores for students at or below that score to that for students above that score. You would use a t-test to test whether two groups differed by a statistically significant amount.
- A more sophisticated analysis would be to compute the regression equation that best describes the relationship between PREview and professionalism scores, and insert the PREview score you set as your threshold for "low" into that equation. The result would be the expected professionalism score associated with PREview score.
    - If the nature your professionalism outcome is dichotomous (binary), such as had an incident or did not have one, your analysis would involve logistic regression.

Do scores from PREview add predictive value – that is, incremental validity, above and beyond other admissions data?

- This analysis of course would require you to have not only PREview scores 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 in this case would be multiple regression, or multiple logistic regression.
- The primary question you would examine is whether the PREview assessment helps predict an outcome after all other admissions data have been considered.
- A secondary question could be whether PREview exam 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 - Correlations

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

MOCK DATA: FOR ILLUSTRATIVE PURPOSES ONLY

	PREview score	Course Grade	Peer Eval: Teamwork	Faculty Eval: Reliability	Turnover
PREview score	1.00				
Course Grade	0.14	1.00			
Peer Eval: Teamwork	0.29	0.09	1.00		
Faculty Eval: Reliability	0.31	0.11	0.68	1.00	
Turnover	0.15	0.19	0.17	0.25	1.00

AAMC

An important consideration when comparing group differences or a examining 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 then look at the size and direction of differences if you were exploring a research question about changing applicant pools, or the size and pattern of the correlations when exploring research questions about the relationship of PREview scores to 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 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 only.)

The PREview score, which is in the first row and column of the table, has low correlations with course grade and turnover. The correlations are, as we might expect, higher for ratings of Peer ratings of Teamwork and with Faculty ratings of Reliability.

Interpret Results – t-test

Research Question	Analysis
How likely is it that someone with a low PREview score will have a professionalism problem in medical school?	• T-test

MOCK DATA: FOR ILLUSTRATIVE PURPOSES ONLY

Professionalism Flag (PF)						
		N	Mean	Std. Dev.	Std. Error Mean	
PF	Low	48	0.062	0.488	0.056	t (123) = .672
	High	77	0.056	0.501	0.072	p=0.503

AAMC

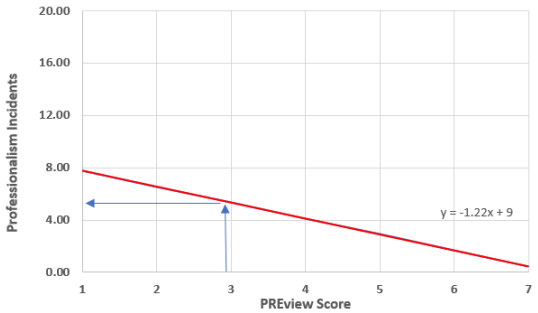
The results of the analyses for the second research question might look something like this. As mentioned, there are at least two ways to look at this question. (Again, bear in mind that these data are made up for illustrative purposes)

In its simplest form, the results of the t-test show that there is no significant difference in the proportion of people with low and high PREview scores who have professionalism flags in their student records.

Interpret Results – Regression

Research Question	Analysis
How likely is it that someone with a low PREview score will have a professionalism problem in medical school?	• Regression analysis

MOCK DATA: FOR ILLUSTRATIVE PURPOSES ONLY



AAMC

The graph shows the more sophisticated analysis we mentioned, with very different findings. The regression equation shows that higher test scores are associated with a lower number of professionalism incidents. Again, this data is for illustrative purposes only.

If a "low" PREview score is defined as a score of 3 or lower, the number of professionalism incidents predicted for someone with a "low" score of 3 is 5 incidents. Conversely, the number of incidents predicted for someone with a score of 7 or higher is zero.

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)	-2203.783	2950.242		-0.747	0.457
Personal Statement	1.58E-07	0	0.064	0.742	0.459
Interview	0.303	0.115	0.228	2.643	0.009
GPA	1.22E-09	0	0.227	2.647	0.009
2(Constant)	-959.448	2764.635		-0.347	0.729
Personal Statement	6.82E-08	0	0.028	0.343	0.732
Interview	0.053	0.121	0.04	0.439	0.662
GPA	1.00E-09	0	0.186	2.31	0.023
PREview	0.468	0.107	0.401	4.388	<.001

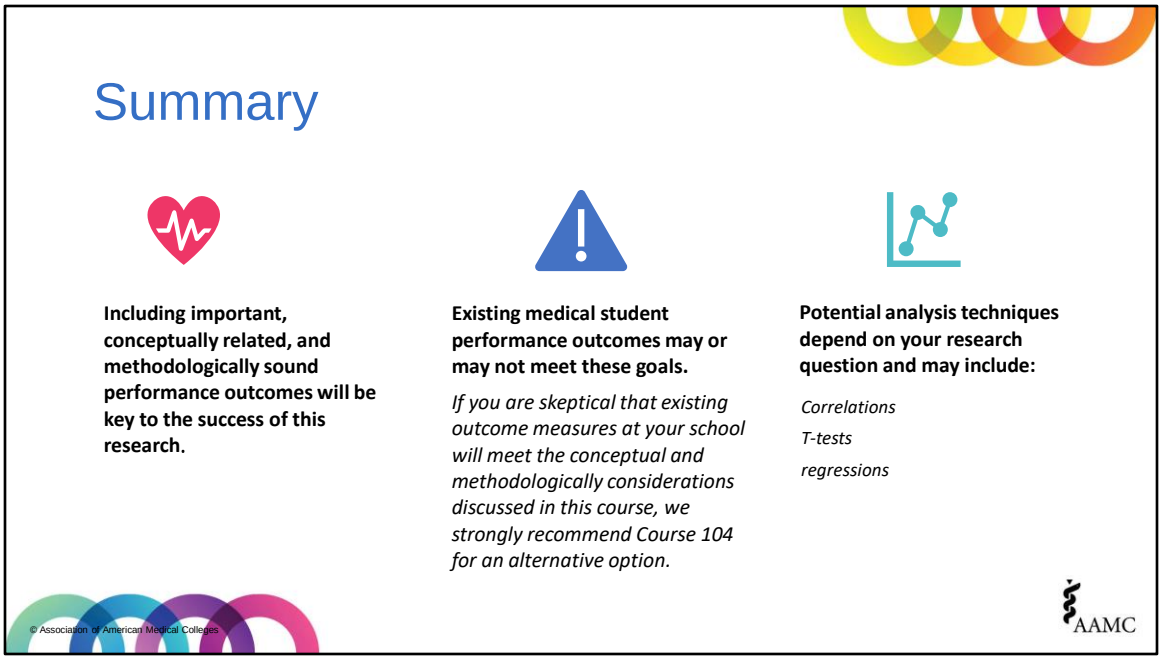
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.331a	0.11	0.088	1.016	0.11	4.972	3	121	0.003
2	.483b	0.233	0.207	0.947	0.123	19.257	1	120	<.001

AAMC

When examining the research question around how much added predictive value PREview scores provide (again results are made up for illustrative purposes only) the interview and GPA do a good job of predicting Dependability, without PREview in the mix.

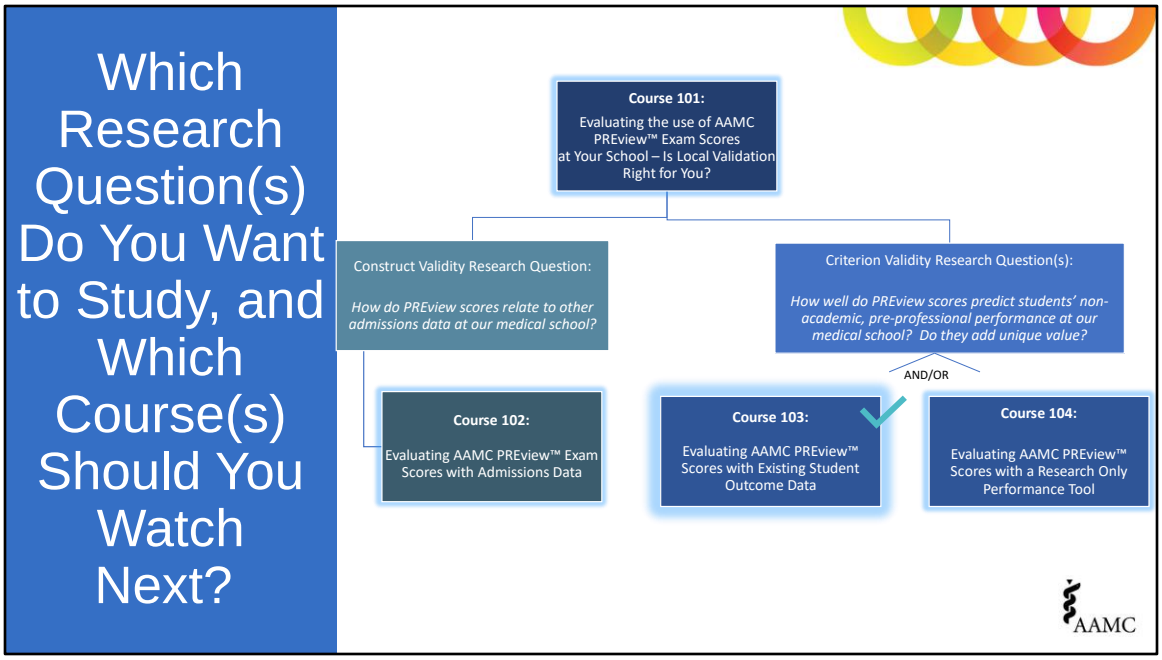
After adding PREview, not only does PREview add to the prediction of Dependability (more than doubling the R square or variance accounted for) it is the best predictor, and in fact, you do a good job of predicting Dependability without Personal Statement or Interview data.

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
- Existing medical student performance outcomes may or may not meet these goals.
- 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, T-tests, and Regression.



Now that you've systematically examined conceptual, methodological, and logistical considerations, is a course-based outcome validation study right for you and your school?

If you are skeptical that existing outcome measures at your school will meet the conceptual and methodological considerations discussed in this course, then Course 104 discusses how to conduct a validity study by collecting your own outcome measures using a research only tool. This tool, developed by the AAMC and free for you to use, can target specific competencies or research questions to help evaluate the use of PREview scores at your school.