



An Overview of IntelliMetric™



Discussion Points:

- ✓ ***What is IntelliMetric™?***
- ✓ ***How is IntelliMetric™ used?***
- ✓ ***How does IntelliMetric™ work?***
- ✓ ***What makes IntelliMetric™ unique?***
- ✓ ***How is a model created?***
- ✓ ***How is IntelliMetric™ trained?***
- ✓ ***How is the quality of a model evaluated?***
- ✓ ***What does the data say about the quality of the scoring?***
- ✓ ***What is a non-scoreable essay?***
- ✓ ***What is The Approximator?***



What is IntelliMetric™?

IntelliMetric is an intelligent automated essay scoring system that:

- ✓ Provides immediate holistic and analytic feedback for extended or short answers
- ✓ May be used as an instructional tool or high stakes assessment
- ✓ Is based on over a decade of research and development with continual innovations and advancements
- ✓ Was first commercially used in 1998 for use in College Course Placement assessments



How is IntelliMetric™ used?

- **Assessment – low through high stakes**
- **Instruction**
- **Short Answer and Extended Response**
- **Multiple languages**
- **With and without human scoring**
- **Grades 4 through Adult**
- **Writing Across the Curriculum**
- **International Implementation**



How does IntelliMetric™ work?

- ✓ Creates a unique scoring solution (model) for each prompt (though non-specific solutions are also available- Approximator)
- ✓ Uses Artificial Intelligence to learn how to score essays
- ✓ Trained with a set of known papers (previously scored by experts)
- ✓ Modeled after traditional expert scoring process

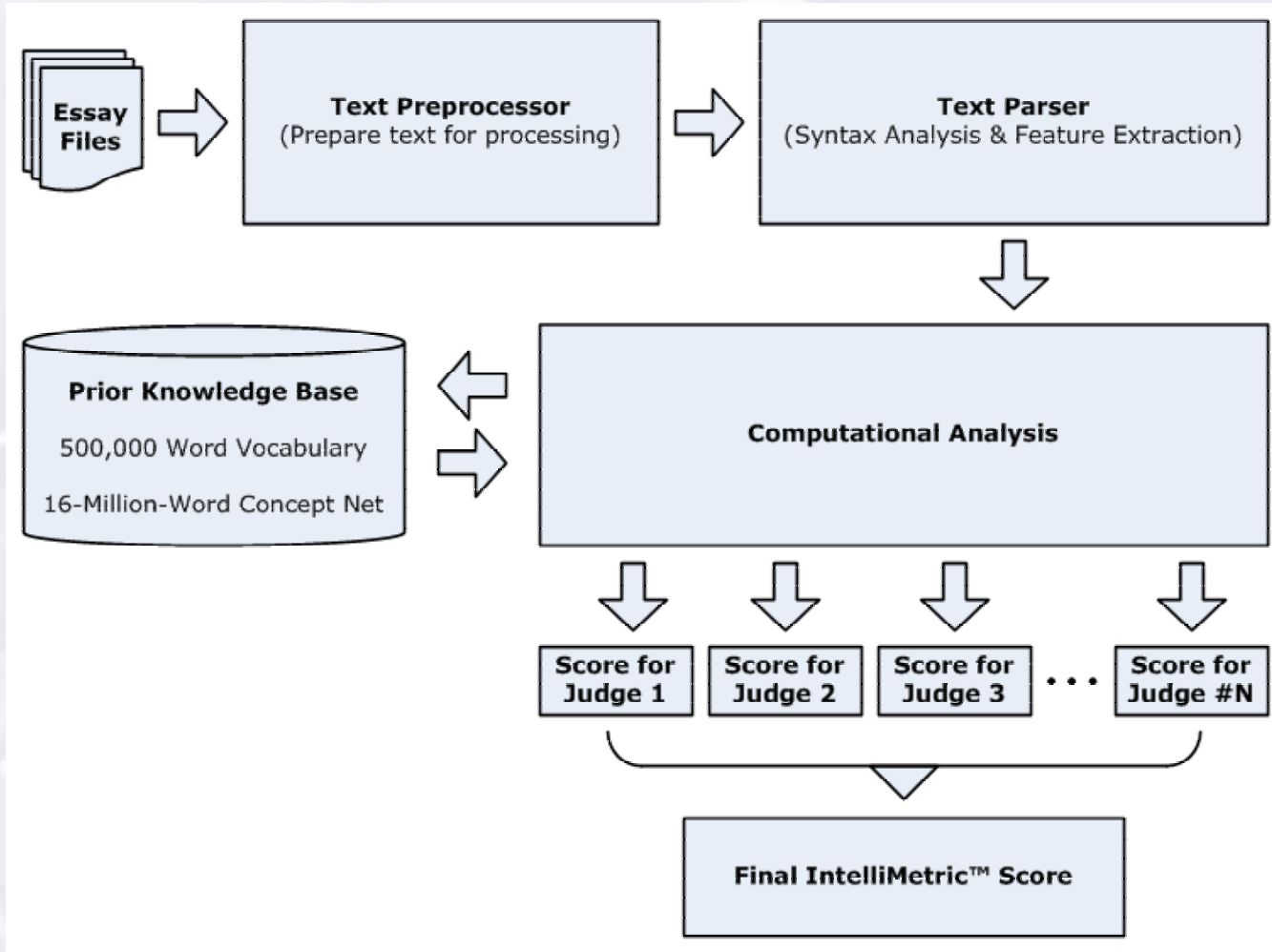


What makes IntelliMetric™ unique?

- ✓ Uses over 400 semantic, syntactic, and discourse-level features
 - ✓ Content and Structure
- ✓ Neurosynthetic approach
- ✓ Learning engine
- ✓ Systemic
- ✓ Inductive
- ✓ Uses Multiple “Judges”
- ⊗ Other automated essay scoring systems use a more narrow approach.



How is an IntelliMetric™ model created?





How is IntelliMetric™ trained?

Prompt specific training mirroring the human scorer calibration process:

- ✓ Training set of approximately 300 representative papers previously scored by expert scorers
 - ✓ Must have definition at tails in order to avoid truncating the scale
 - ✓ Must be consistently scored- garbage in, garbage out
 - ✓ Roughly normal distributions provide strongest models

- ✓ Validation set kept blind



Continued.....How is IntelliMetric™ trained?

- ✓ IntelliMetric™ infers rubric from training set
IntelliMetric™ is inductive rather than rule based
- ✓ Model is created and used to score validation set
- ✓ Performance data is reviewed (comparing human score and IM score for validation set)
Agreement Rates
Pearson correlation (measure of linear relationship)



How do we evaluate the quality of a model?

- ✓ Confusion Matrix (Crosstabulation of scores)
- ✓ Classification Analysis
- ✓ Pearson Correlation
- ✓ Agreement rates- Exact, Adjacent, Discrepant

Example of a Crosstabulation in which Humans and IntelliMetric are in perfect agreement:

	C1	C2	C3	C4	C5	C6
H1	5	0	0	0	0	0
H2	0	10	0	0	0	0
H3	0	0	10	0	0	0
H4	0	0	0	10	0	0
H5	0	0	0	0	10	0
H6	0	0	0	0	0	5



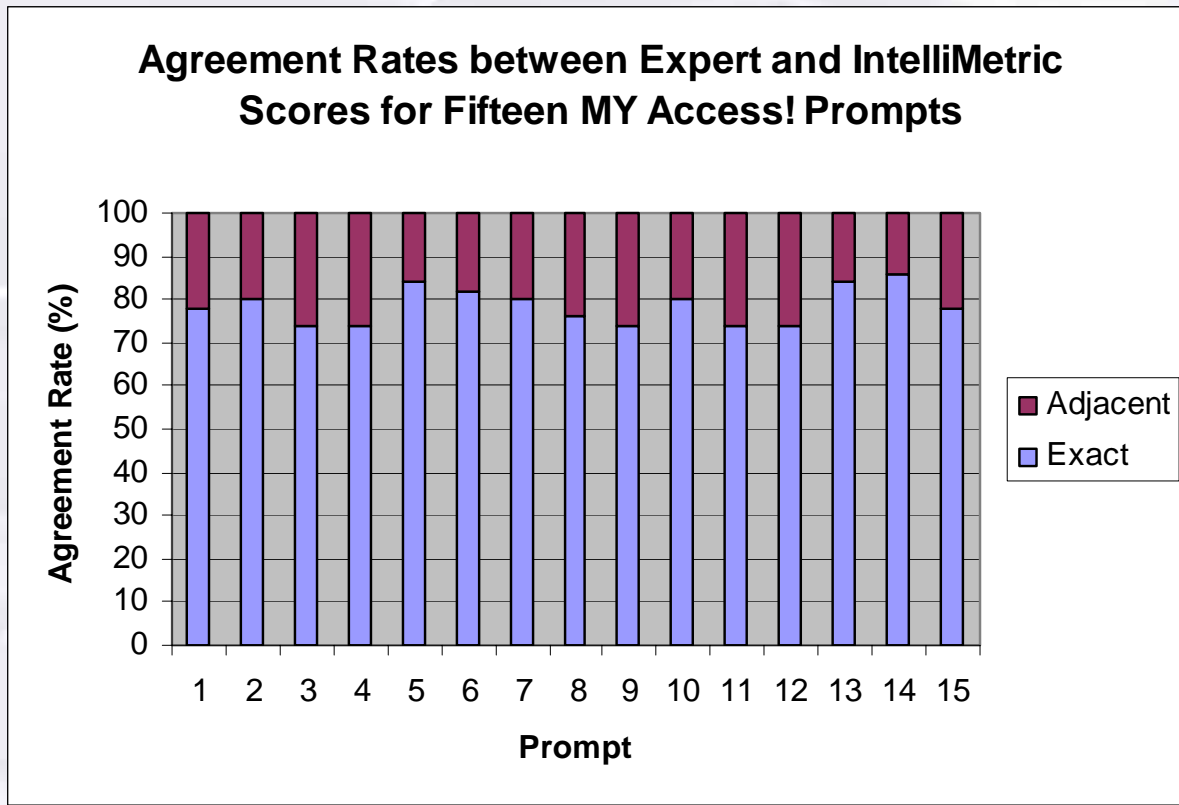
What does the data say about the quality of the scoring?

Hundreds of studies provide support that IntelliMetric™:

- ✓ Agrees with expert scoring, often exceeding the performance of expert scorers
- ✓ Accurately scores open-ended responses across a variety of grade levels, subject areas and contexts
- ✓ Shows a strong relationship with other measures of the same writing construct
- ✓ Shows stable results across samples



What does the data say about the quality of the scoring?





What does the data say about the quality of the scoring?

- ✓ IM scores agree with human scores more often than two skilled human raters will agree with each other.
- ✓ Whether human or IM scoring, domain scoring is less accurate than holistic scoring
- ✓ 6-point rubrics seem to do the best
- ✓ As a result of the way IM “learns,” IM is able to model different kinds of rubrics on different scales with different domains
- ✓ IM is perfectly internally consistent. It will consistently score the same essay the same way every time. This is not true of human scorers.



What is a non-scoreable essay?

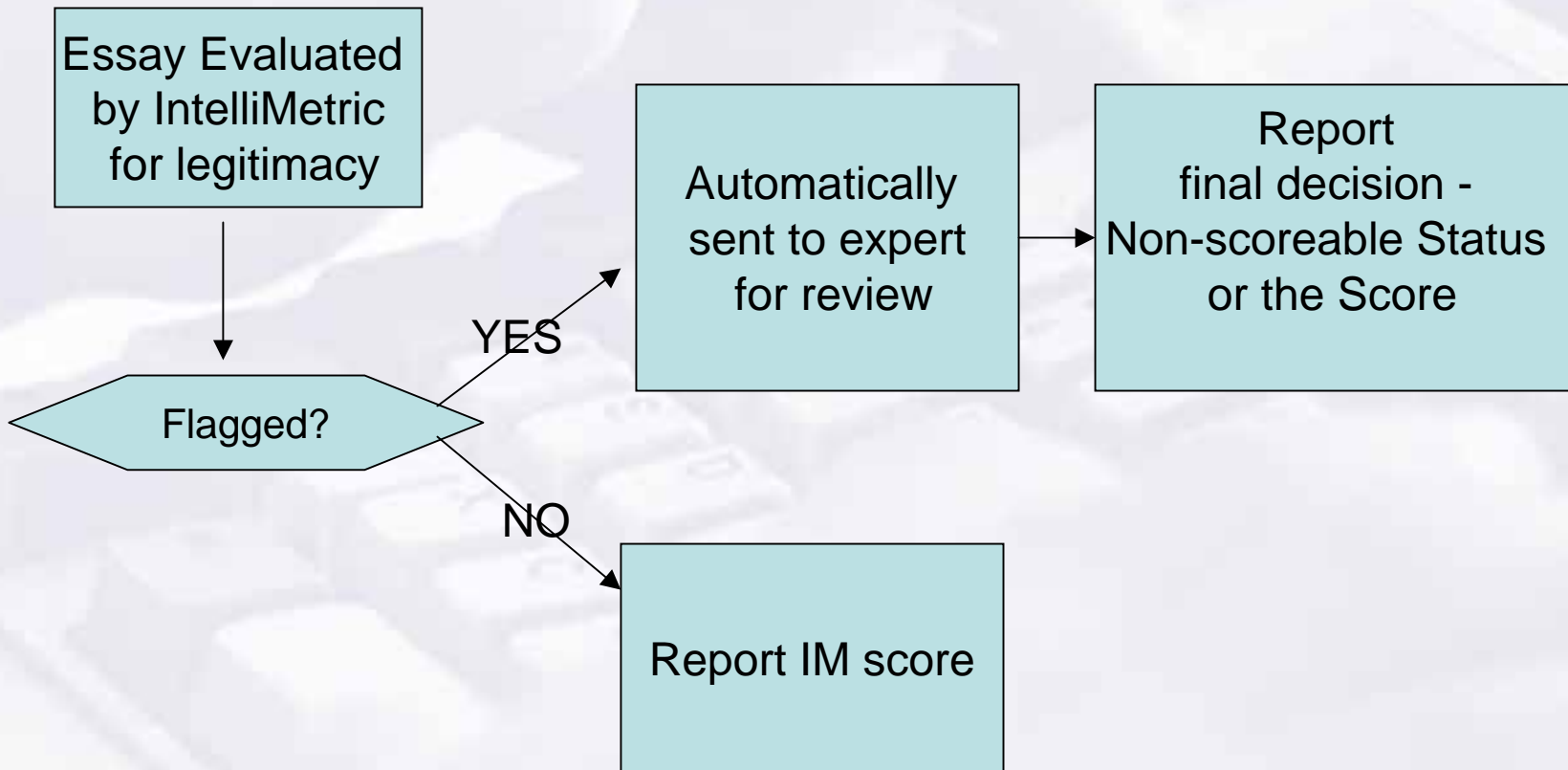
✓ Each time IntelliMetric attempts to score an essay, IntelliMetric also goes through a series of checks to determine if the essay is legitimate. Some possible causes for concern:

- ✓ Off topic
- ✓ Off task
- ✓ In another language
- ✓ Too short to evaluate
- ✓ Insufficient development
- ✓ Bad Syntax (Gobbledygook)
- ✓ Repetitious (pieces of essay copied throughout or overuse of certain phrases)
- ✓ Inappropriate
- ✓ Copy of prompt
- ✓ Copy of passage or related materials provided



What happens to non-scoreable essays?

In MY Access!®, each non-scoreable is reviewed by an expert human scorer to make the final decision.





What is The Approximator™?

- ✓The Approximator uses the same training and modeling process as IntelliMetric, yet the Approximator is trained on a pool of prompts rather than a unique prompt. This provides a more “generic” model for scoring a wide variety of prompts.
- ✓No non-scoreable checks for Approximator
- ✓No domain scores are provided