



Innovations in Testing 2024

Know thy enemy... items. And the variants, too!

BETTER TOGETHER
Embrace change. Share solutions.



Russ Smith, Ph.D.

Chief Innovation
Officer & Senior
Psychometrician



Natalie Flacco

Product Specialist II
Exam Development



Nick Respecki

Senior Manager
of Examination



About PMI...

We empower project professionals around the globe to excel in project management practices through our growing global community, knowledge sharing, and best-in-class certifications—driving positive change in organizations and communities.

About PMI

Diverse portfolio of exams leading to credentials in various aspects of project management

**Certified Associate in
Project Management
(CAPM)®**

**Project Management
Professional (PMP)®**

**Program Management
Professional (PgMP)®**

**PMI Risk Management
Professional (PMI-RMP)®**

**PMI Construction
Professional Project
Management Professional
(PMI-CP)™**

About NCARB...

The National Council of Architectural Registration Boards protects the public health, safety, and welfare by leading the regulation of the practice of architecture through the development and application of standards for licensure and credentialing of architects.

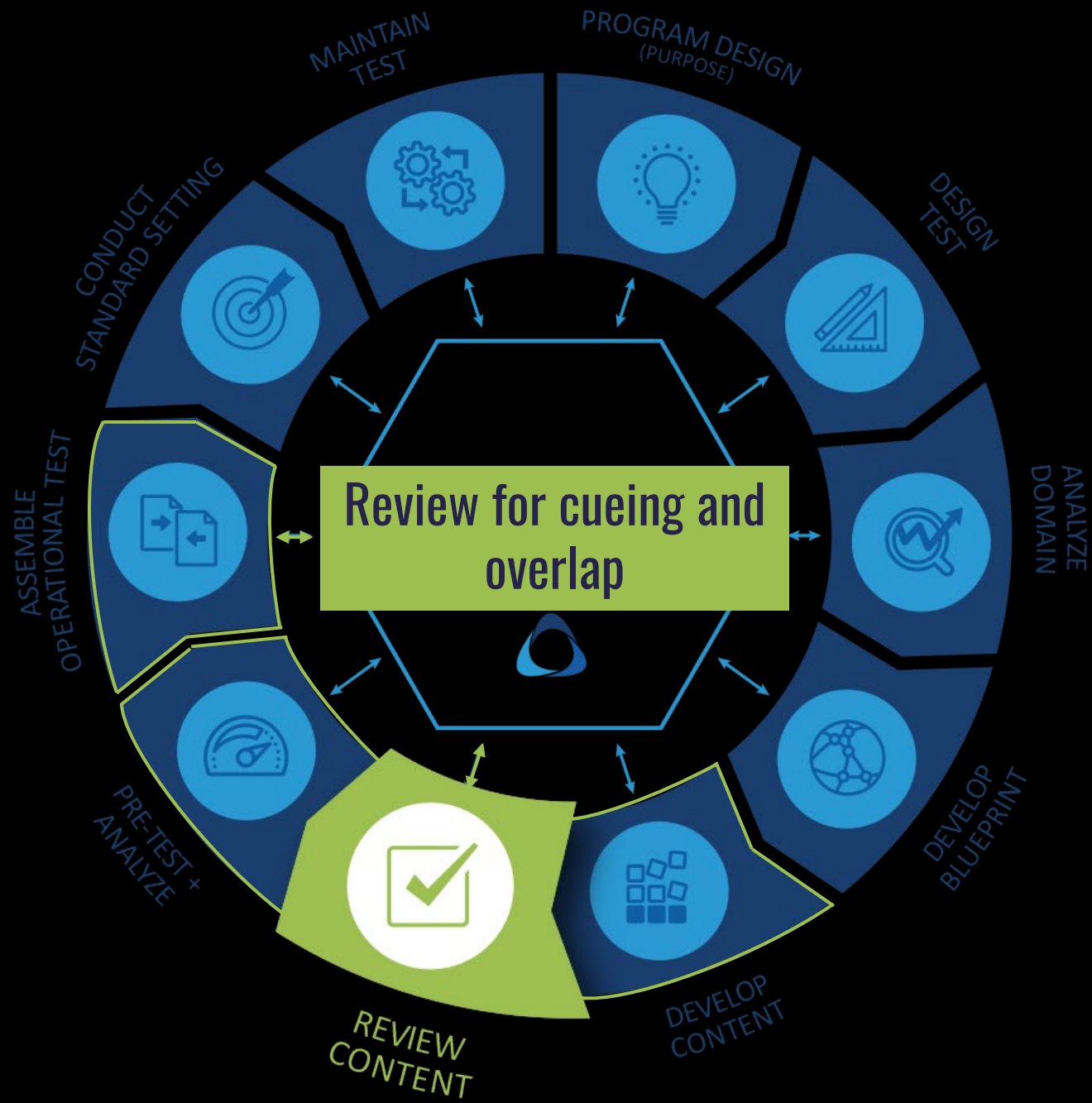
Problem: Comparing items for cueing and overlap

1,000 item bank is...

... nearly half a million comparisons!

@ 30 seconds a pair...

... it would take 2 years!



Problem: Comparing items for cueing and overlap

150 item form is...

over 11,000 comparisons!

@ 30 seconds a pair...

it would take 2 weeks

... per form!

Literature review

Becker, K.A. and Kao, S. (2022)

- * Identifying Enemy Item Pairs using Natural Language Processing

Peng, F. (2020)

- * Automatic enemy item detection using natural language processing

Weir, J. B. (2019)

- * Enemy item detection using data mining methods

Model	No. of predicted duplicates	Accuracy
TF-IDF(unigram stem)	476	79.33
TF-IDF(unigram key)	307	51.17
TF-IDF(unigram stem+key)	487	81.17
TF-IDF(uni,bi,trigram stem)	480	80.00
TF-IDF(uni, bi, trigram key)	312	52.00
TF-IDF(uni, bi, trigram stem+key)	507	84.50
TF-IDF+LDA(unigram stem)	376	62.67
TF-IDF+LDA(unigram key)	212	35.33
TF-IDF+LDA(unigram stem+key)	343	57.17
TF-IDF+LDA(uni,bi,trigram stem)	380	63.33
TF-IDF+LDA(uni, bi, trigram key)	216	36.00
TF-IDF+LDA(uni, bi, trigram stem+key)	346	57.67

Automated Enemy Detection

Process items in bank

Frequency count for each word

Calculate Similarity

Flag and review pairs with high similarity

Processing text: **Tokenize**, **Stop Words**, **Lemmatize**

An accepted deadline for a project approaches. However, the project manager realizes only 75% percent of the work has been completed. The project manager then issues a change request. What should the change request authorize?

Corrective action based on causes

accept deadline approach however
realize percent work complete issue
change request change request
authorize corrective action based
cause

Automated Enemy Detection

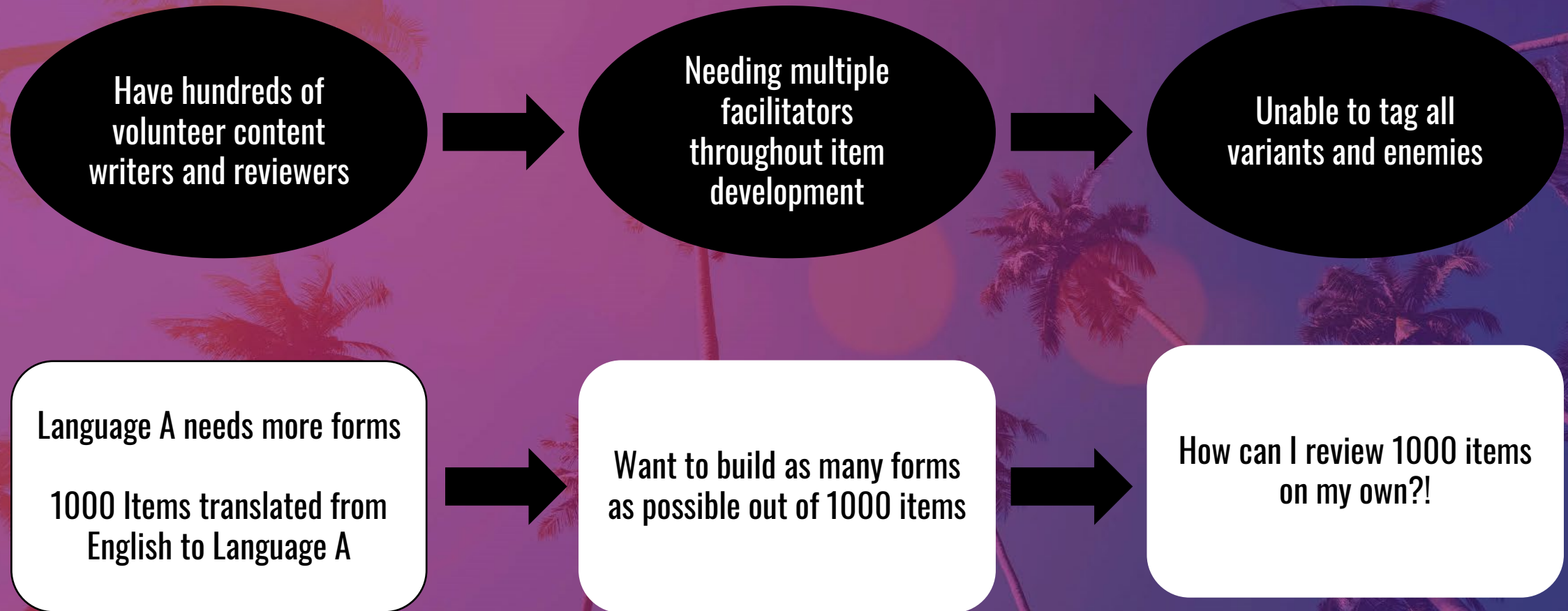
Process items in bank

Frequency count for each word

Calculate Similarity

Flag and review pairs with high similarity

PMI – Background



Results: PMI

First pass with the tool and key word search

Item 1	Item 2	Already marked as	Should be	Caught in	Stem to Stem	Key to Key	Options to Options	Stem 1 to Key 2	Stem 2 to Key 1
Item 1	Item 2	Variants	variant	Key Word Search					
538959	812332	No	Yes	Yes	0.07	0.95	0.4	0	0.01
538959	539019	No	Yes	Yes	0	0.1	0.03	0	0.02
668473	812326	No	Yes	Yes	0	0.22	0.02	0	0
747724	800503	No	Yes	Yes	0	0	0	0	0
747724	812248	No	Yes	Yes	0	0.01	0	0.01	0
747751	800751	No	Yes	Yes	0	1	0.32	0	0

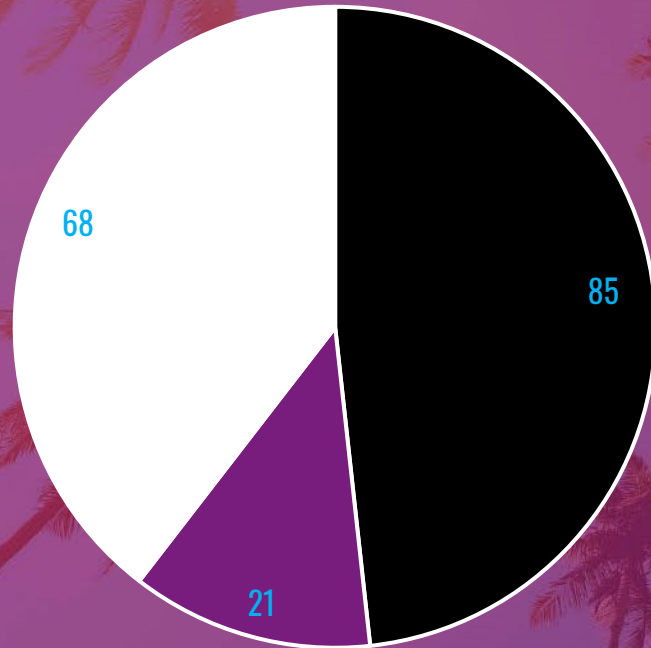
Results: PMI

Next step

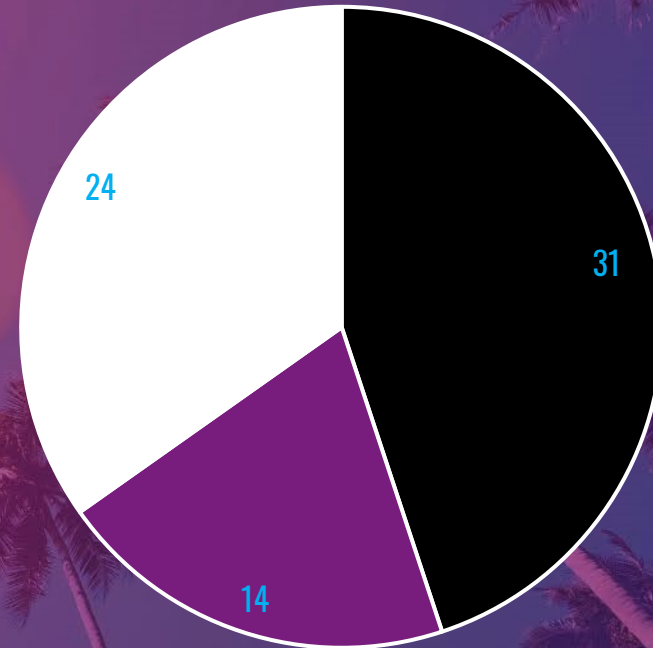
- * Terminology list of key words
- * Include synonyms
 - 🌐 Impediment, obstacle, blockers
- * Removing terms
 - 🌐 Project, project manager, product, project lead, etc.
- * Run it again

Results: PMI

Variant Pairs



Additional Pairs By Type



■ Pairs Caught ONLY by Me ■ Pairs Caught by both Tool and Me ■ Pairs Caught ONLY by Tool

■ Stem to Stem ■ Key to Key ■ Cueing

Results: PMI

Next Steps

- * Add more words to the terminology list
- * Make more terms “stop words”
- * Continue Iterating

Lessons Learned

- * The role meaning plays in our content
 -  Need more than the just the analysis of the words used

Results (overlap): NCARB

By the numbers...

- * **458 items** reviewed for overlap
- * **83 item pairs** identified for overlap
- * **69 of 83 item pairs** contained overlap or warranted further SME review

- * **Volunteer SMEs – better with small item sets**
- * **Automated Enemy Detection – better with large item sets**

Results (cueing): NCARB

By the numbers...

- * **458 items** reviewed for cueing
- * **82 item pairs** identified for cueing
- * **21 of 82 item pairs** contained cueing or warranted further SME review
- * Overall, not as successful as overlap detection but has potential
- * Stop words can reduce false flags (consultant, engineer, contractor, etc.)
- * **Volunteer SMEs – better with small item sets**
- * **Automated Enemy Detection – better with large item sets**

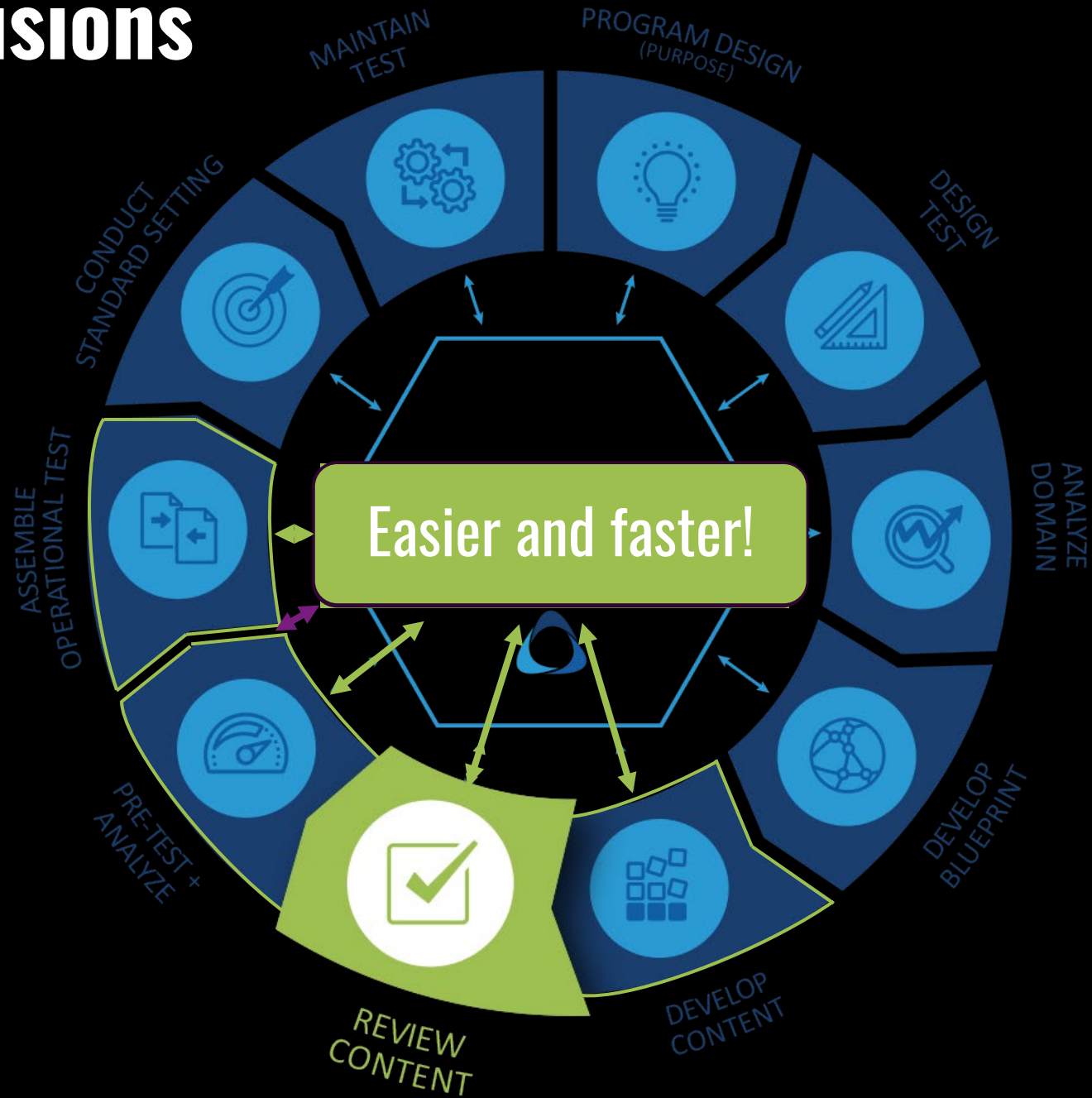
On a large item set, your method did a better job of identifying potential item overlap than our committee members.

I'm excited to see this aid exam development programs throughout the industry.

You're on to something here, Russ!



Conclusions



What's the point?

- ✦ NOT to replace humans
- ✦ Make humans more efficient and effective
- ✦ Reduce time & cost



ATP
GLOBAL



BETTER TOGETHER

Embrace change. Share solutions.