SLIDES ONLY
NO SCRIPT PROVIDED

OPA
OFFICE OF PEOPLE ANALYTICS

Development of a Complex Reasoning (CR) Test

Katherine Klein
Human Resources Research Organization (HumRRO)

Briefing presented to the DACMPT
August 16, 2023

Background

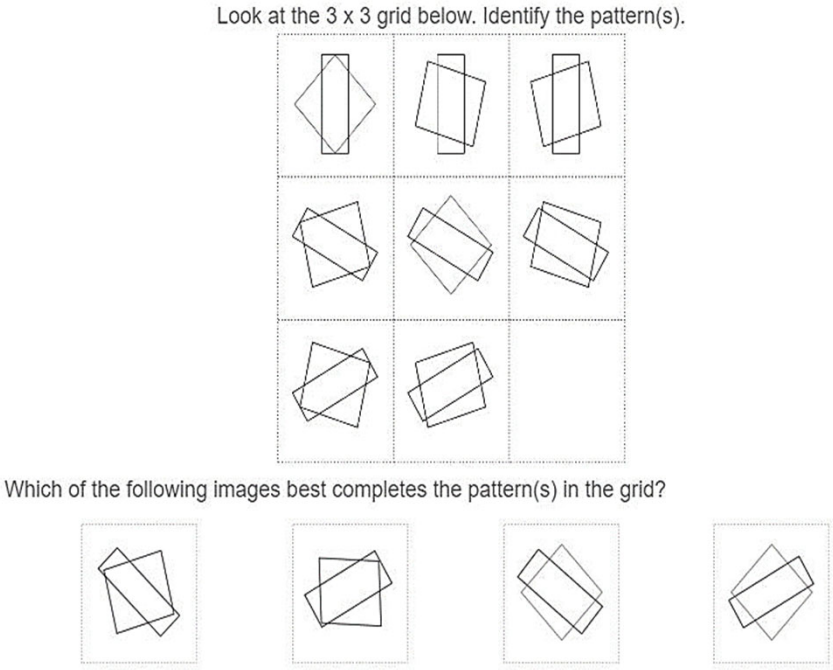
- ***What is complex reasoning?***
  - Non-verbal reasoning; ability to analyze visual information and to solve problems using visual reasoning
- ***Why a complex reasoning test?***
  - Fluid intelligence has been found to be a strong predictor of training and job success
    - Complex (non-verbal) reasoning is one element of fluid intelligence
    - ASVAB Review Panel (2006) recommended that DoD consider adding tests of fluid intelligence to balance the ASVAB's composition (between fluid and crystalized intelligence)
  - Potential benefits to the ASVAB testing program
    - Improved prediction of training and job success in military jobs
    - Lower susceptibility to test compromise
    - Less adverse impact; increased qualification rates for non-native and non-heritage English speakers

Current Development Effort

Objective: Develop a complex reasoning (non-verbal) testing system to generate items for potential inclusion on ASVAB

- Employ non-proprietary Automated Item Generation (AIG) capability
  - Improve item development efficiency
  - Reduce or eliminate field-testing requirements
- Generate items with targeted properties
  - Items similar to Raven's Progressive Matrices (RPM) items
  - Items at appropriate difficulty for qualifying military applicants into jobs of varying complexity

Sample Transformation Item



- **Transformation item features**
  - Types of shapes
  - Orientation of shape(s)
  - Size of shape(s)
  - Number of shape(s)
  - Line weighting on shape(s)
- **Direction(s) of transformations**
  - Vertical
  - Horizontal
  - Diagonal

OPA
OFFICE OF PEOPLE ANALYTICS

Complex Reasoning (CR) Test Development Program

Line of Effort (LOE)	Progress
LOE 1: Develop, Pilot, and Evaluate Initial CR Capability	COMPLETED
LOE 2a: Develop an Improved CR Item Generation Tool	COMPLETED
LOE 2b: Pilot and Evaluate Refined CR Capability	IN PROGRESS, projected completion September 2023
LOE 3: Develop Operational CR Test Form(s) and Future R&D/Maintenance Plans	IN PROGRESS, projected completion January 2024

OPA
OFFICE OF PEOPLE ANALYTICS

LOE 1: CR Pilot Study 1 Recommendations

- Use transformation items only
- Use four response options, no “none of these are correct” (NOTAC) option
- Refine item difficulty model and item selection to ensure appropriate level of difficulty and minimize group score differences by race-ethnicity, where feasible

LOE 2b: CR Pilot Study 2 Overview

Objective

- Collect data on refined pool of CR items representative of the population of CR items with a participant sample representative of military applicants
  - Collect sufficient data to evaluate group score differences on CR items and forms among a sample representative of military applicants
- Results will be used to:
  - Develop CR form(s) for operational implementation on ASVAB platform to support Computational Thinking requirement and related research
  - Select pool of experimental CR items for potential inclusion with operational CR form(s)
  - Inform future R&D and test maintenance plans for CR

LOE 2b: CR Pilot Study 2 Overview (continued)

Design and Measures

- 24 CR items, 3 static forms, same 24 items on each form but in a different fixed order (spiraled by estimated difficulty)
- Pre- and post-test questionnaire
  - Demographics, perceived difficulty of items, test-taking experience
- Two CR attention check items + items measuring insufficient effort responding

LOE 2b: CR Pilot Study 2 Overview (continued)

Sample

- Non-military sample representative of military applicants, ages 18–35, U.S. citizen, HS degree/GED/<1 year of college
- Targeted *N* = 2,600 participants
  - ~866 participants per form

Method

- Administered on Qualtrics platform
- Participants randomly assigned to one CR form
- No fixed time limit; record time to completion
- Desktop or laptop only

OPA
OFFICE OF PEOPLE ANALYTICS

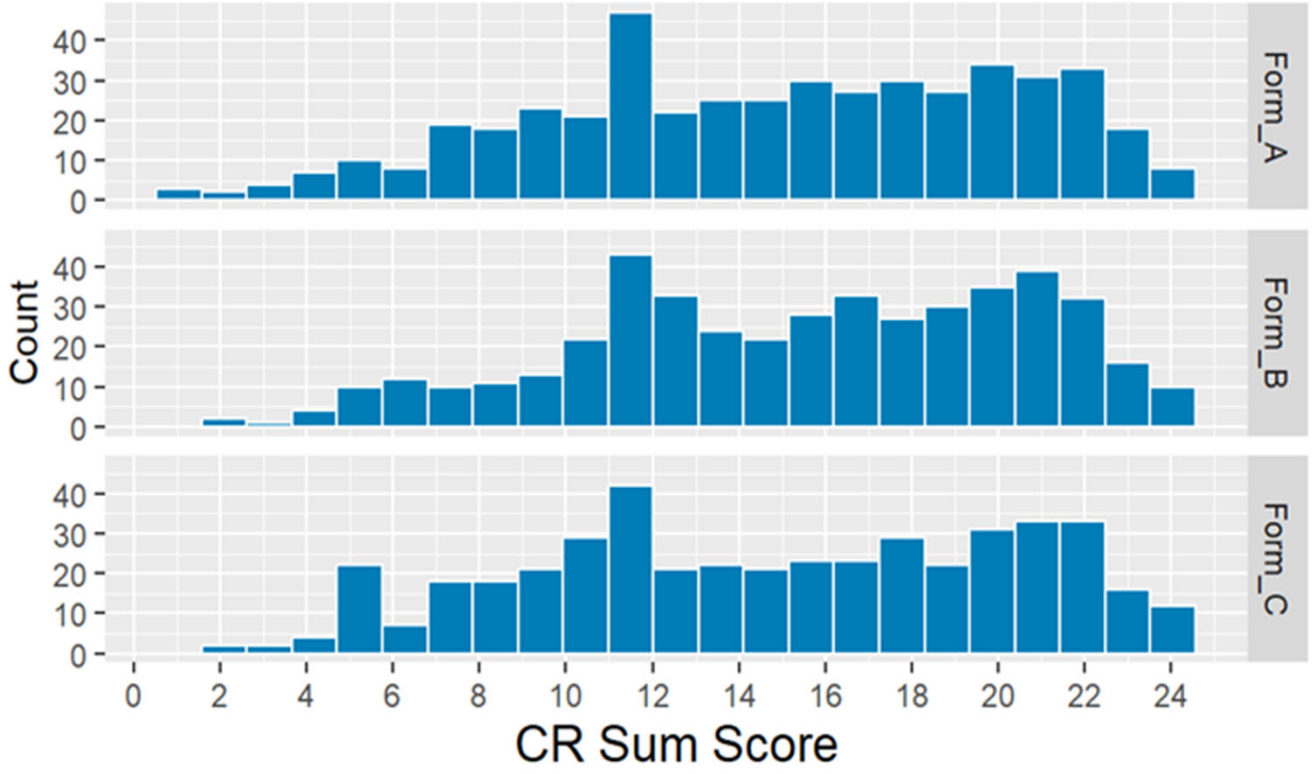
LOE 2b: CR Pilot Study 2 Data Collection Summary (as of 20 June 2023)

Group	Pilot 2 (as of 20 June 2023)				Pilot 1 (Summer 2022)
	Form A	Form B	Form C	Combined (All Forms)	Transform Only, 4 Options
Total	472	457	451	**1,380**	188
Female	297	288	294	**879**	113
Asian	39	43	31	**113**	10
Black	127	125	100	**352**	28
Hispanic	191	170	208	**569**	42

Note. Pilot 1 numbers reflect the subset of participants that match the sample frame for Pilot 2 (ages 18–35, U.S. citizen, HS degree/GED/<1 year of college) and who completed Pilot 1 in 30 minutes or less.

OPA
OFFICE OF PEOPLE ANALYTICS

LOE 2b: CR Pilot Study 2 Test Scores Summary (as of 20 June 2023)



Form A
n = 472, *M* = 14.92, *SD* = 5.53, avg *p* = .62

Form B
n = 457, *M* = 15.59, *SD* = 5.18, avg *p* = .65

Form C
n = 451, *M* = 14.86, *SD* = 5.64, avg *p* = .62

CR Pilot 1 (Summer 2022)
n = 188, *M* = 13.89, *SD* = 5.12, avg *p* = .58

OPA
OFFICE OF PEOPLE ANALYTICS

LOE 3: Develop CR Test Form(s) + R&D/Maintenance Plans

Objectives

- Develop CR test form(s) for operational implementation on ASVAB platform to support Computational Thinking requirement and related research (Computational Thinking composite score)
  - Four (4) static CR forms for operational implementation; same items (<= 24 items) on each form but in a different fixed presentation order (spiraled by difficulty)
  - Supplemented with pool of experimental items for future implementation (e.g., overlong forms) or R&D
- Develop future R&D and test maintenance plans for CR

Deliverables (September 2023)

- Four (4) static CR forms for operational implementation
- Pool of experimental CR items for future implementation
- Info for scoring CR items and generating a CR test score

OPA
OFFICE OF PEOPLE ANALYTICS

Points of Discussion

- Operational Implementation of CR Test Forms and Scores
- Future R&D

OPA
OFFICE OF PEOPLE ANALYTICS

Acknowledgments

Matthew Brown, *HumRRO*

Scott Oppler, *HumRRO*

Mike Ingerick, *HumRRO*

Sergio Marquez, *HumRRO*

Sachi Phillips, *HumRRO*

Furong Gao, *HumRRO*

Mary Pommerich, *DTAC*

Tia Fechter, *DTAC*

Jeff Harber, *DTAC*

OPA
OFFICE OF PEOPLE ANALYTICS

Thank you!

For more information please contact:

Katherine Klein
KKlein@HumRRO.org
651.370.210

OPA
OFFICE OF PEOPLE ANALYTICS

PHASE 1 PILOT: SUMMING UP (ALL PARTICIPANTS, *N* = 3,491)

Metric	Transform Only 8	Transform Only 3 + NOTAC	Transform Only 4	Transform Only 4 + NOTAC	Transform + Logic Grouped	Transform + Logic Scrambled
Unidimensionality	Yes	No	Yes	No	No	No
Reliability	α = .87 SEM = 1.98 *avg* CITC = .46	α = .78 SEM = 1.97 *avg* CITC = .31	α = .85 SEM = 1.99 *avg* CITC = .42	α = .83 SEM = 1.96 *avg* CITC = .37	α = .75 SEM = 2.10 *avg* CITC = .32	α = .75 SEM = 2.04 *avg* CITC = .30
Observed Difficulty	*M* = 12.29 *SD* = 5.48 avg *p* = .51	*M* = 13.29 *SD* = 4.21 avg *p* = .54	*M* = 15.00 *SD* = 5.15 avg *p* = .63	*M* = 11.96 *SD* = 4.76 avg *p* = .50	*M* = 10.32 *SD* = 4.19 avg *p* = .43	*M* = 9.15 *SD* = 4.08 avg *p* = .38
Group Score Differences	F-M *d* = .21 B-W *d* = -.39 H-W *d* = -.17 A-W *d* = .36	F-M *d* = .31 B-W *d* = -.31 H-W *d* = -.15 A-W *d* = .17	F-M *d* = .22 B-W *d* = -.58 H-W *d* = -.21 A-W *d* = .00	F-M *d* = .16 B-W *d* = -.20 H-W *d* = -.19 A-W *d* = .12	F-M *d* = .10 B-W *d* = -.34 H-W *d* = -.22 A-W *d* = .59	F-M *d* = .22 B-W *d* = -.08 H-W *d* = .04 A-W *d* = .24
Completion Time (30 minutes or less)	*M* = 12.74 *SD* = 5.86	*M* = 10.88 *SD* = 4.94	*M* = 11.36 *SD* = 5.14	*M* = 12.39 *SD* = 5.79	*M* = 13.54 *SD* = 6.18	*M* = 13.24 *SD* = 6.36
Perceived Difficulty	*M* = 3.92 *SD* = .95	*M* = 3.89 *SD* = .92	*M* = 3.98 *SD* = .89	*M* = 3.90 *SD* = .95	*M* = 3.50 *SD* = .96	*M* = 3.37 *SD* = 1.00

Note. NOTAC = None of these are correct. SEM = Standard error of measurement. CITC = Corrected item-total correlation.

PHASE 1 PILOT: SUMMING UP (COMPLETED ≤ 30 MINUTES WITH HS DEGREE/GED/< 1 YR OF COLLEGE, *N* = 1,200)

Metric	Transform Only 8	Transform Only 3 + NOTAC	Transform Only 4	Transform Only 4 + NOTAC	Transform + Logic Grouped	Transform + Logic Scrambled
Unidimensionality	--	--	--	--	--	--
Reliability	α = .88 SEM = 1.78 *avg* CITC = .39	α = .79 SEM = 1.78 *avg* CITC = .28	α = .86 SEM = 1.92 *avg* CITC = .40	α = .84 SEM = 1.76 *avg* CITC = .37	α = .67 SEM = 2.11 *avg* CITC = .27	α = .69 SEM = 1.98 *avg* CITC = .26
Observed Difficulty	*M* = 11.60 *SD* = 5.15 avg *p* = .48	*M* = 12.56 *SD* = 3.89 avg *p* = .52	*M* = 13.89 *SD* = 5.12 avg *p* = .58	*M* = 11.66 *SD* = 4.39 avg *p* = .49	*M* = 9.71 *SD* = 3.68 avg *p* = .40	*M* = 8.27 *SD* = 3.56 avg *p* = .35
Group Score Differences	F-M *d* = .20	F-M *d* = .26	F-M *d* = .26	F-M *d* = .18	F-M *d* = -.02	F-M *d* = .34
Completion Time (30 minutes or less)	*M* = 12.32 *SD* = 5.70	*M* = 10.59 *SD* = 4.91	*M* = 11.34 *SD* = 5.41	*M* = 12.04 *SD* = 5.51	*M* = 13.24 *SD* = 6.16	*M* = 12.79 *SD* = 6.21
Perceived Difficulty	*M* = 3.83 *SD* = .97	*M* = 3.79 *SD* = .93	*M* = 3.86 *SD* = .91	*M* = 3.83 *SD* = .96	*M* = 3.41 *SD* = .97	*M* = 3.26 *SD* = 1.00

Note. NOTAC = None of these are correct. SEM = Standard error of measurement. CITC = Corrected item-total correlation.