

Appendixes

Appendix 1. *Inter-rater Reliability Indexes among Three Post-Units Assessments*

Workbook	Number of raters	Inter-rater correlation coefficient	Kappa	Percent of agreement
Unit 1				
1	3	0.57	0.49	0.65
2	3	0.51	0.42	0.61
3	2	0.65	0.47	0.68
4	2	0.88	0.62	0.63
5	2	0.75	0.58	0.69
6	2	0.74	0.52	0.69
		68%	52 %	66 %
Unit 2				
1	4	0.78	0.69	0.71
2	4	0.76	0.65	0.74
3	4	0.7	0.59	0.75
4	4	0.75	0.6	0.76
		75 %	63 %	74 %
Unit 3				
1	3	0.84	0.56	0.71
2	4	0.76	0.56	0.7
3	4	0.77	0.55	0.7
4	4	0.75	0.58	0.71
5	4	0.73	0.55	0.7
6	4	0.66	0.43	0.6
7	4	0.59	0.44	0.59
8	4	0.65	0.46	0.62
9	3	0.64	0.41	0.58
10	4	0.67	0.53	0.63
		69 %	50.70%	64 %
Overall Average		71 %	54 %	67%

Note: Workbook indicates each round of student assessment scoring, which includes both using model and developing model.

Appendix 2. Content Validity of the Cognitive Flexibility Items: Expert Review.

Cognitive Flexibility Expert Review Scoring Guidance

Thank you for your willingness to provide your expertise for our assessment tasks for measuring cognitive flexibility. Please respond to the three questions below. Each question refers to linked related assessment tasks.

Note: For the purpose of this review, we are defining the elements of the DCIs as “separate meanings the same DCI takes as listed in the NGSS for explaining a phenomenon”. For example,

LS4.B: Natural Selection

- Sometimes the differences in characteristics between individuals of the same species provide advantages in surviving, finding mates, and reproducing.

could be used to describe advantages in finding mates *or* reproducing.

Multiple Pathways Example: To understand what is meant by multiple pathways, a student could explain how a particular toy, a ball, moves using ideas of forces of gravity, magnetism, or friction. All pathways might be acceptable, and then, if they are, the task would be considered eliciting multiple pathways.

For the tasks, please open the relevant documents in the NGSS in third grade.

1. Tasks set 1 Squirrels

Read the entire assessment. Please respond to the question with respect to the questions: 2, 2b, 2c, 3a.

See: Relevant documents in the NGSS [3-LS1](#) [3-LS2](#) [3-LS3](#) [3-LS4](#)

Questions regarding assessments

1. To what degree are the assessment items ill-structured and suitable for measuring cognitive flexibility through eliciting demand for more than one pathway for an application of different DCIs or different elements of the same DCI's in third grade?
 - a. The assessment items offer potential for multiple pathways (i.e., use of more than two DCI's and/or elements of one DCI's).
 - b. The assessment items offer not more than 2 pathways.
 - c. There is only one pathway toward an acceptable response to the items.

Answer _____

Rationale

2. Tasks set 2 toys

Read the entire assessment. Please respond to the question with respect to the questions: 8, 9, 10 and 11

See: Relevant documents in the NGSS [3-PS2 3-5-ETS1](#)

Questions regarding assessments:

2. To what degree are the assessment items ill-structured and suitable for measuring cognitive flexibility through eliciting demand for more than one pathway for an application of different DCIs or different elements of the same DCI's in third grade?

- a. The assessment items offer potential for multiple pathways (i.e., use of more than two DCI's and/or elements of one DCIs.
- b. The assessment items offer not more than 2 pathways.
- c. There is only one pathway toward an acceptable response to the items.

Answer _____

Rationale

3. Tasks set 3 birds 3-LS1 3-LS2 3-LS3 3-LS4

Read the entire assessment. Please respond to the question with respect to the questions: 1,2, 3 and 4

See: Relevant documents in the NGSS

Questions regarding assessments:

3. To what degree are the assessment items ill-structured and suitable for measuring cognitive flexibility through eliciting demand for more than one pathway for an application of different DCIs or different elements of the same DCI's in third grade?

- a. The assessment items offer potential for multiple pathways (i.e., use of more than two DCI's and/or elements of one DCIs.
- b. The assessment items offer not more than 2 pathways.
- c. There is only one pathway toward an acceptable response to the items.

Answer _____

Rationale

Appendix 3. Results of Invariance Tests of a Second-Order Factor Model of Cognitive Flexibility CFA

By group	Second-Order CFA Measurement Invariance Model	χ (df) ^a	$\Delta \chi$ (df)	CFI	Δ CFI	RMSEA
Gender	M1. Configural invariance model	87.25(48)		0.96		0.046
	M2. Metric invariance of the first-order factors Loading	98.15(57)	10.9 (9)	0.95	0.01	0.043
	M3. Metric invariance of the first- and second-order factors Loading	99.76(59)	1.61(2)	0.95	0.00	0.042
	M4. Scalar invariance of the first-order factor intercept	107.59(65)	7.83(6)	0.95	0.00	0.041
	M5. Scalar invariance of the first- and second-order factor intercept	118.20(68)	10.61(3)	0.94	0.01	0.048
ELL status	M1. Configural invariance model	74.32(48)		0.97		0.037
	M2. Metric invariance of the first-order factors Loading	96.13(57)	21.81 (9)	0.96	0.01	0.042
	M3. Metric invariance of the first- and second-order factors Loading	98.58(59)	2.45 (2)	0.95	0.01	0.041
	M4. Scalar invariance of the first-order factor intercept ^b	135.86(65)	37.28(6)*	0.92	0.03	0.053
	M5. Scalar invariance of the first- and second-order factor intercept ^c	181.95(68)	46.09(3)*	0.87	0.05	0.065
Economic Difficulty	M1. Configural invariance model	93.32(48)		0.9		0.047
	M2. Metric invariance of the first-order factors Loading	101.47(57)	8.15 (9)	0.89	0.01	0.049
	M3. Metric invariance of the first- and second-order factors Loading	103.28(59)	1.81(2)	0.88	0.01	0.05
	M4. Scalar invariance of the first-order factor intercept	n/a		n/a		n/a
	M5. Scalar invariance of the first- and second-order factor intercept	n/a		n/a		n/a
Non-white/ White	M1. Configural invariance model	80.19(48)		0.96		0.041
	M2. Metric invariance of the first-order factors Loading	96.88(57)	16.69 (9)	0.95	0.01	0.042
	M3. Metric invariance of the first- and second-order factors Loading	98.14(59)	1.26(2)	0.95	0.00	0.041
	M4. Scalar invariance of the first-order factor intercept ^b	135.86(65)	37.72(6)*	0.91	0.04	0.053
	M5. Scalar invariance of the first- and second-order factor intercept ^c	181.59(68)	45.75(3)*	0.87	0.04	0.065

Note. * significant at $p < 0.05$ as estimated by DIFFTEST procedure in Mplus.

- Scaled χ^2 difference is a difference between -2log-likelihood with maximum likelihood robust estimation; df – degrees of freedom;
- Scalar invariance of the first-order factor intercept (4) was assessed by fixing the first-order factor loadings to be equal across groups. The model fit reaches significantly when comparing to the above model condition (3), which implied that the levels of the first-order latent variables correspond to the different levels of the manifest variables across groups.
- Scalar invariance of the first-order and second-order factor intercept (5) was assessed by fixing the first-order and second-order factor loadings to be equal across groups. The model fit reaches significantly when comparing to the above model condition (4), which implied that the levels of the second-order latent variables correspond to the different levels of the manifest variables across groups.