

Summary of Research on the STAGES Developmental Model

Terri O'Fallon and Tom Murray

ESRAD 2021 Conference

(slides at
<https://www.slideshare.net/perspegrity5/presentations>,
video at: <https://youtu.be/UloSrzuKP0A>)

STAGES research published papers

- O'Fallon, T., Polissar, N., Neradilek, M. B., & Murray, T. (2020). The validation of a new scoring method for assessing ego development based on three dimensions of language. *Elsevier Heliyon*, 6(3)
- *Integral-Review.org*: 18 articles - Special Issue on the STAGES model. Vol. 16 No. 1 April 2020.
- Murray, T. & O'Fallon, T. (2020). **Summary of STAGES Validation Research** (above v. 16 no. 1)

The STAGES Matrix: Three Parameters or “Questions”

PP	Question 1: Is the object of awareness Concrete, Subtle, or MetAware?	Question 2: Is the experience Individual or Collective?	Question 3: Is the experience Receptive, Active, Reciprocal, or Interpenetrative?	STAGE NAME
	TIER	SOCIAL PREFERENCE	LEARNING SEQUENCE	
1.0	Concrete	Individual	Receptive	Impulsive
1.5	Concrete	Individual	Active	Egocentric
2.0	Concrete	Collective	Reciprocal	Rule Oriented
2.5	Concrete	Collective	Interpenetrative	Conformist
3.0	Subtle	Individual	Receptive	Expert
3.5	Subtle	Individual	Active	Achiever
4.0	Subtle	Collective	Reciprocal	Pluralist
4.5	Subtle	Collective	Interpenetrative	Strategist
5.0	MetAware	Individual	Receptive	Construct Aware
5.5	MetAware	Individual	Active	Transpersonal
6.0	MetAware	Collective	Reciprocal	Universal
6.5	MetAware	Collective	Interpenetrative	Illumined

Research studies summary

1. *STAGES vs. SCTi-MAP Replication Study*
2. Inter-Rater Studies at the item scale
3. *Longitudinal Analysis*
4. Analysis of Late Stage Patterns
5. Rasch Analysis of item and test strength
6. Investigating the validity of the Ogive Aggregation Method
7. *Specialty Inventories – custom SCTs*
8. *Studies of Children's Inventories*
9. *Vocabulary Analysis*
10. Face Validity: STAGES Applications

Fire Hose Presentation!

- In about 15 minutes we want to give you a quick overview of new studies related to validating the STAGES developmental model.
- After these slides are slides with another version of the entire presentation with more detail on each item.
- (All the info is also in the IR Special issue referenced above)

1. STAGES vs. SCTi-MAP Replication

- Q: Is STAGES assessment method measuring the same construct as the SCTi-MAP – up through Strategist (4.5)?
- > A Convergent validity / replicability Study
- > 73 inventories, 4 scorers (stratified sampling)
- **RESULT: “excellent” match using Cohen’s Kappa**
- (**Does not mean exact match!)

	N	Weighted Kappa	
		Stages 2.0 & 2.5 separated	Stages 2.0 & 2.5 combined
<i>Scorer 1 (most experienced)</i>	73*	0.95	0.94
<i>Scorer 2</i>	48	0.82	0.84
<i>Scorer 3</i>	48	0.79	0.81
<i>Scorer 4</i>	50	0.88	0.87
Mean (All)	73	0.86	0.87
Mean (excluding Scorer 1)	73	0.83	0.84

* All 73 inventories were scored by Scorer 1. Each inventory was scored by 2 of the other 3 scorers.

Table 3. Inter-rate Reliability of Tier 3 (stages <5 combined into one category)

	Kappa for Set A (n=71)	Kappa for Set B (n=51)
All scorers	0.65	0.68
Scorer 1 excluded	0.56	0.63

2. Inter-Rater Reliability – new study at **item level**

- **Q: How good is scorer training at the item level?**
- (Ave 98% IRR had been established at the survey level)
- **RESULT: 5 scorers - ITEM LEVEL: IRR accuracy Ave 93%**

Table 6. Rater accuracy aggregated over 10 scorings (of the 36-item test) for each rater.

Scorer	Average Acc	Min Acc
1	0.88	0.72
4	0.93	0.83
3	0.91	0.86
2	0.94	0.89
5	0.97	0.92
Average	0.93	

3. Longitudinal Study (Prior vs. next scores)

- Strong proof of “development” is longitudinal increase
- Q: How often do successive assessments (> 6 months apart) increase or stay the same (vs. decrease)?
- RESULTS:
 - N=143: Increase: 50%; Same: 39%; Decrease: 11%
 - Metaware only: Increase: 66%; Same: 30%; Decrease: 4%

Table 4. Longitudinal: prior scores vs. amount of change to next score.

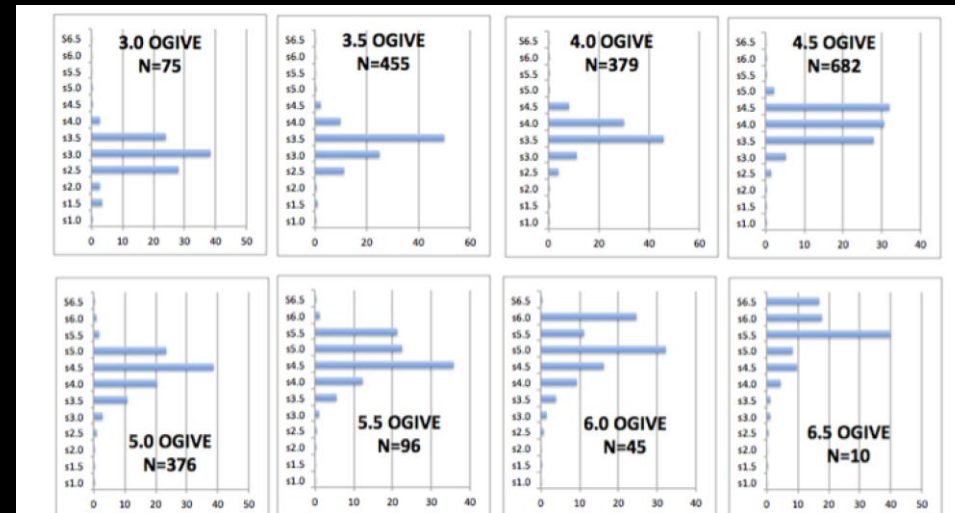
Change to Next Score								
Prior Score	-2	-1	-0.5	0	0.5	1	1.5	Grand Total
3.5				3	6	2	1	12
4			2	6	8	3		19
4.5		1	3	21	27	6		58
5			4	15	12	1		32
5.5	2	1	2	9	5			19
6			1	1	1			3
Grand Total	2	2	12	55	59	12	1	143

Table 5. Longitudinal data when the new score is in the MetAware tier.

Prior Score	Score Change					Grand Total
	-0.5	0	0.5	1	1.5	
3.5					1	1
4				3		3
4.5			27	6		33
5		15	12	1		28
5.5	2	9	5			16
6	1	1	1			3
Grand Total	3	25	45	10	1	84

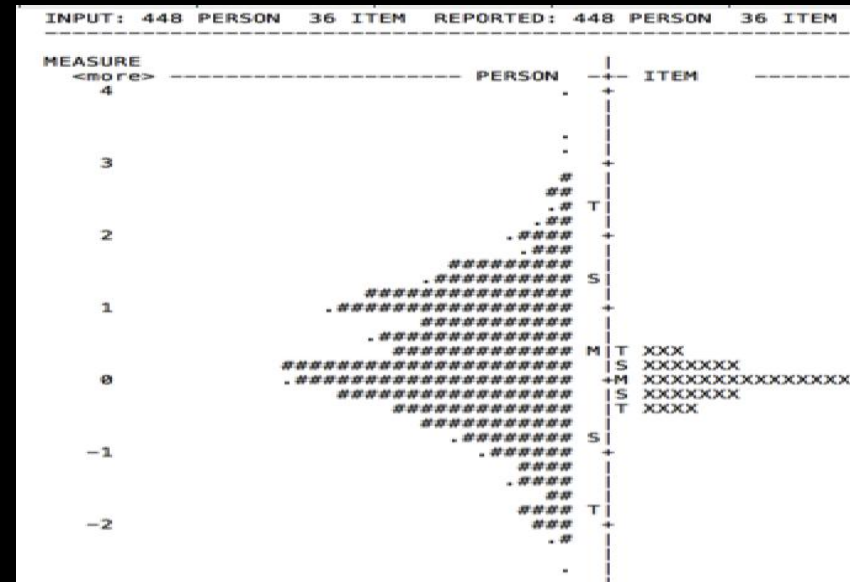
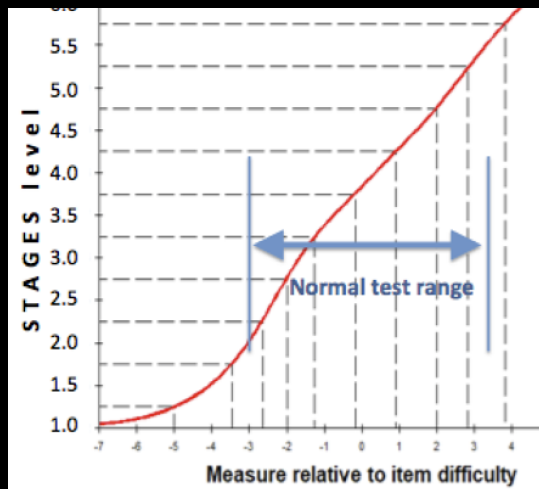
4. Metaware challenge

- Q: Are those scoring in the Metaware tier merely displaying “esoteric spiritual language”
- or do they also **have sufficient 4.5/Strategist** scores, showing they include that complexity?
- RESULT: **metaware scores show complex strategist** completions.



5. Rasch Analysis at item level

- Q: How does STAGES validate through modern Rasch (and “item response theory”) validation?
- RESULTS:
 - Test looks strong
 - Levels are evenly spaced



6. Drawbacks of the cutoff aggregation method

- **PROBLEM:** there are many issues with the “ogive” cutoff method used in the Loevinger tradition, as used to aggregate the 36 stems to get the total survey score.
- **SOLUTION:**
 - “**Core Score**” – based on **average of the top 6 scores**
 - Gives continuous (i.e. fractional) results
 - Includes a “Bottom” (or shadow) score
 - Does not confuse early development with “shadow crash” items
 - Allows studying children (see below)

7. Specialty Inventories

- STAGES allows for easy creation of **new stems and topic-focused inventories**.
- Q: “What does development look like in domain X?”
- Results: **Alpha validations look excellent**

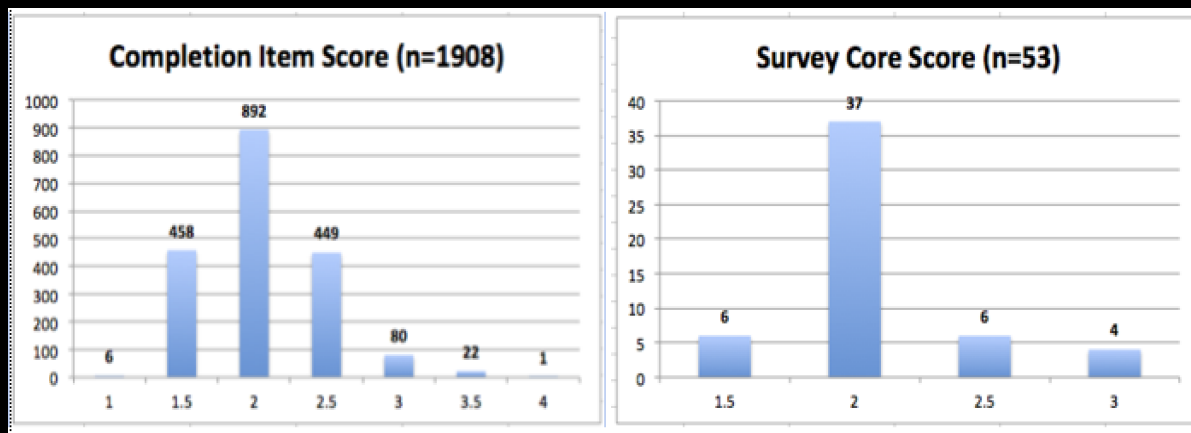
Table 2. Cronbach's alpha internal consistency values for specialty protocols.

Domain	Cronbach's alpha					
	Leadership	Love	Education	Climate Change	Psych	Children
N in study	32	20	20	32	17	53
New stems (of 36)	6	6	6	6	5	11
α All 36 stems	0.97	0.95	0.95	0.96	0.92	0.88
α General prot, stems only	0.96	0.94	0.94	0.95	0.90	0.85
α New stems only	0.85	0.82	0.83	0.82	0.80	0.63

Education	validated
Love	validated
Climate Change	validated
Dementia	validated
Hope	validated
Leadership/Business	validated
children	validated
Relationships	validated
Money	validated
Psychology	validated
Mormon (LDS) spirituality	data-gathered
Health Care for Chronic Illness	data-gathering
Alternative (incl.Chinese) Medicine	data-gathering
Resilience	data-gathering
Anomalous experiences	
Ethics	
Gender	

8. Children's Development - ages 4-13

- First study of its kind to study ego development in children
- n=53,; verbal completions transcribed
- **Acceptable internal validity** (less than adult tests)
- Full analysis in progress
- Made possible by our new Aggregation method



9. Vocabulary Analysis (preliminary)

- Ongoing analysis of word frequencies at each level.
 - Thematic analysis (e.g. “love”)
 - Track drift over time
 - Fine-tune the scoring procedure



10. “Face Validity” of STAGES

PhD research using STAGES

- Lynam, A. - **Teaching** and Mentoring in Higher Education
- Hochachka, G. - Importance of human interiority in transformative approaches to **climate change**
- Mantler, N. - **Women’s** Authentic **Leadership** Development
- Churchill, J. - Indo-Tibetan **Contemplative** Essence Psychology
- Braks, A. - Transformational Executive **Coaching**
- Buckley, L. - **Hope** Examined Through a Developmental Stage Perspective*
- Reynolds, E. - Next-Stage **Organizations**: A Transdisciplinary Case Study
- Attebery, J. - **Sustainability** Pedagogy for Incarcerated Adult Learners
- Parker, J. - STAGES of **Organizational Development** *
- Hurst, R. - Ego Development and Self Initiated **Reflection**
- Aguocha, F. – **Relational Leadership** and Ego Development
- (* in progress)

Masters Theses and other applications

- Miller, J. - Importance of **Reflective Practice** in Self-Knowledge.
- John Kesler's "**Integral Polarity Practice**" (IPP)
- Marj Britt, Senior Minister Emeritus in the Unity Church, work on "the soul's journey through life [as a] **cosmic and human love story**."
- Pacific integral's GTC (**Generating Transformative Change**) cohort program (Geoff Fitch and Abigail Lynam)
- Translations to **German Language** (in process)
- Hilbert, M. - **Susceptibility to Fake News** and media manipulation (beginning)

RECAP: STAGES Research

1. *STAGES vs. SCTi-MAP Replication Study*
2. Inter-Rater Studies at the item scale
3. *Longitudinal Analysis*
4. Analysis of Late Stage Patterns
5. Rasch Analysis of item and test strength
6. Investigating the validity of the Ogive Aggregation Method
7. *Specialty Inventories – custom SCTs*
8. *Studies of Children's Inventories*
9. *Vocabulary Analysis*
10. Face Validity: STAGES Applications

END

- (Of Fire-hose presentation)

Learn More!

- StagesInternational.com
- Slides at
<https://www.slideshare.net/perspegrity5/presentations>
- Planned: follow-up conversation next week, same time—email tommurray.us@gmail.com

DETAILED SLIDES

- The following are details of the slides shown in the Fire-hose presentation above

1. STAGES vs. SCTi-MAP Replication

- Convergent validity:
 - STAGES (scoring method: based on 3 parameters)
vs.
 - SCTi-MAP (scoring method: similar to original
Loevinger method, mainly based on examples)
- 73 inventories, 4 scorers (stratified sampling)
- 8 Levels: Concrete & Subtle Tiers (1.0-4.5 Strat.)

The STAGES Matrix: Three Parameters or “Questions”

PP	Question 1: Is the object of awareness Concrete, Subtle, or MetAware?	Question 2: Is the experience Individual or Collective?	Question 3: Is the experience Receptive, Active, Reciprocal, or Interpenetrative?	STAGE NAME
	TIER	SOCIAL PREFERENCE	LEARNING SEQUENCE	
1.0	Concrete	Individual	Receptive	Impulsive
1.5	Concrete	Individual	Active	Egocentric
2.0	Concrete	Collective	Reciprocal	Rule Oriented
2.5	Concrete	Collective	Interpenetrative	Conformist
3.0	Subtle	Individual	Receptive	Expert
3.5	Subtle	Individual	Active	Achiever
4.0	Subtle	Collective	Reciprocal	Pluralist
4.5	Subtle	Collective	Interpenetrative	Strategist
5.0	MetAware	Individual	Receptive	Construct Aware
5.5	MetAware	Individual	Active	Transpersonal
6.0	MetAware	Collective	Reciprocal	Universal
6.5	MetAware	Collective	Interpenetrative	Illumined

Tier 1-2 Study (1.0 to 4.5 Strategist)

	N	Weighted Kappa	
		Stages 2.0 &2.5 <u>separated</u>	Stages 2.0 &2.5 <u>combined</u>
<i>Scorer 1 (most experienced)</i>	73*	0.95	0.94
<i>Scorer 2</i>	48	0.82	0.84
<i>Scorer 3</i>	48	0.79	0.81
<i>Scorer 4</i>	50	0.88	0.87
<i>Mean (All)</i>	73	0.86	0.87
<i>Mean (excluding Scorer 1)</i>	73	0.83	0.84

* All 73 inventories were scored by Scorer 1. Each inventory was scored by 2 of the other 3 scorers.

- Tier1-2 study (up to 4.5), there was
"substantial" or "excellent" agreement

Metaware Tier - Reliability

Table 3. Inter-rate Reliability of Tier 3 (stages <5 combined into one category)

	Kappa for Set A (n=71)	Kappa for Set B (n=51)
All scorers	0.65	0.68
Scorer 1 excluded	0.56	0.63

Cohen's Kappa - Interpretation

Cohen's Kappa statistic (κ)	Strength of agreement
< 0.00	Poor
0.00–0.20	Slight
0.20–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

Table 1. Agreement details for STAGES vs. Cook-Greuter/Loevinger systems.

(CG/L)*	1.0 (2)	1.5 (2/3)	2.0-2.5 (3)	3.0 (3/4)	3.5 (4)	4.0 (4/5)	4.5 (5)	Total	<u>Agrmnt</u>	Agree +/- 1
STAGES										
1.0	4	0	0	0	0	0	0	4	100.0%	100.0%
1.5	5	5	3	2	0	0	0	15	33.3%	86.7%
2.0-2.5	3	5	35	9	1	0	0	53	66.0%	92.5%
3.0	0	2	9	21	1	0	0	33	63.6%	93.9%
3.5	0	0	1	4	28	7	3	43	65.1%	90.7%
4.0	0	0	0	0	3	19	6	28	67.9%	100.0%
4.5	0	0	2	0	3	9	23	37	62.2%	86.5%
5.0	0	0	0	0	0	1	4	5		
Total	12	12	50	36	36	36	36	218		
<u>Agrmnt</u>	33.3%	41.7%	70.0%	58.3%	77.8 %	52.8%	63.9%			
Agree +/- 1	75.0%	83.3%	94.0%	94.4%	88.9 %	97.2%	91.7%			

2.0–2.5, 3.5, 4.0 slight *higher*; 3.0, 4.5 slight *lower*

2. Inter-Rater Reliability – new study at **item level**

- **Industry standard:** IRR accuracy of 85% at survey level
- **Survey level:** 5 scorers, IRR accuracy at the was Ave. **98%**

Table 6. Rater accuracy aggregated over 10 scorings (of the 36-item test) for each rater.

Scorer	Average Acc	Min Acc
1	0.88	0.72
4	0.93	0.83
3	0.91	0.86
2	0.94	0.89
5	0.97	0.92
Average	0.93	

- **ITEM LEVEL:** IRR accuracy Ave **93%**

3. Longitudinal Study

Prior vs. next scores

Table 3. Longitudinal: prior vs. next scores.

		Next Score						
Prior Score	3.5	4	4.5	5	5.5	6	6.5	Grand Total
3.5	3	6	2	1				12
4	2	6	8	3				19
4.5	1	3	21	27	6			58
5			4	15	12	1		32
5.5	2		1	2	9	5		19
6					1	1	1	3
Grand Total	8	15	36	48	28	7	1	143

Longitudinal: by Changes

Table 4. Longitudinal: prior scores vs. amount of change to next score.

Prior Score	Change to Next Score							Grand Total
	-2	-1	-0.5	0	0.5	1	1.5	
3.5				3	6	2	1	12
4			2	6	8	3		19
4.5		1	3	21	27	6		58
5			4	15	12	1		32
5.5	2	1	2	9	5			19
6			1	1	1			3
Grand Total	2	2	12	55	59	12	1	143

- Increase: 50%; Same: 39%; Decrease: 11%
- (Removing data from GTC graduates, still only 17% decreased)

Longitudinal Study - Metaware tier

Table 5. Longitudinal data when the new score is in the MetAware tier.

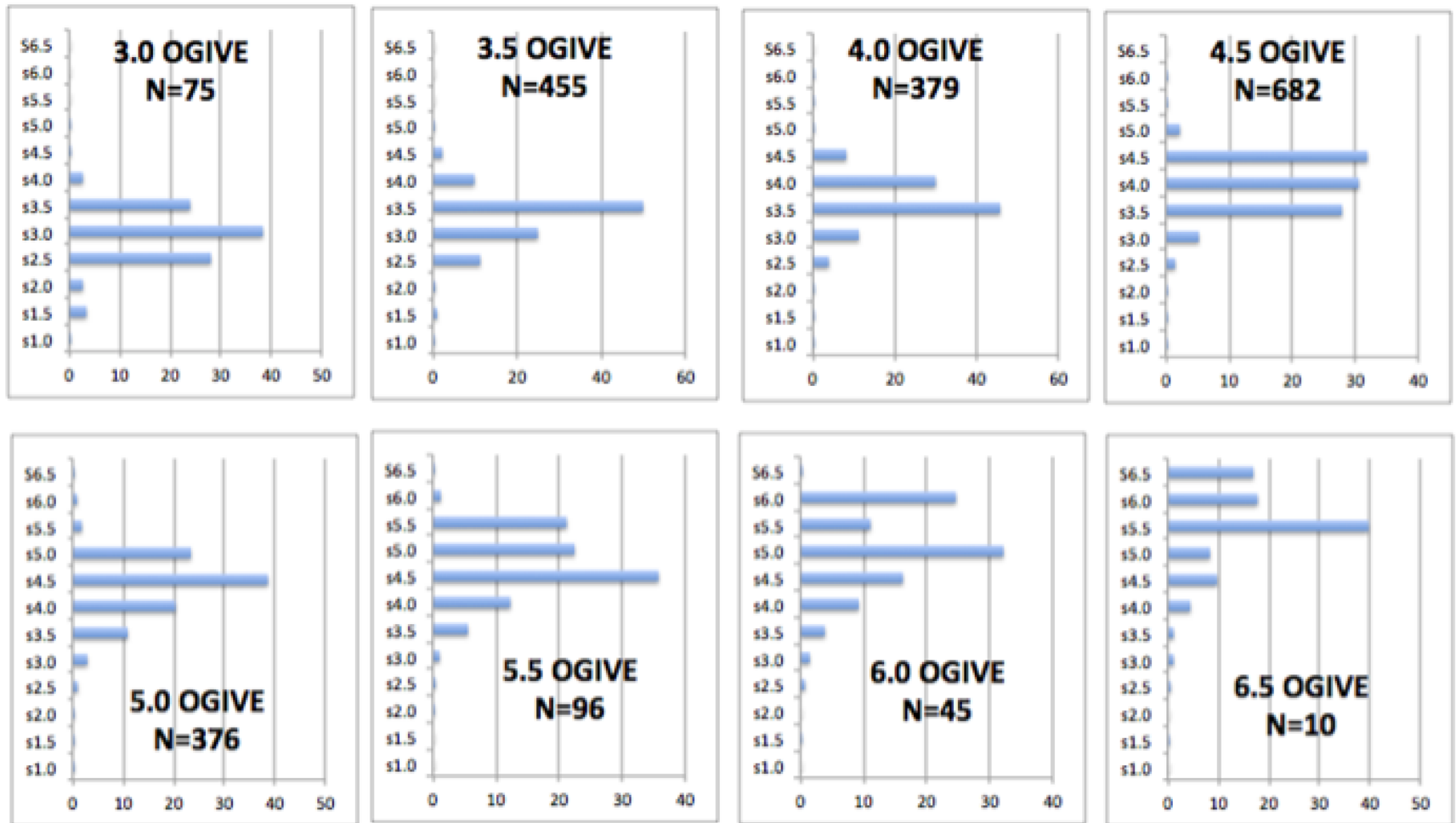
Prior Score	Score Change					Grand Total
	-0.5	0	0.5	1	1.5	
3.5					1	1
4				3		3
4.5			27	6		33
5		15	12	1		28
5.5	2	9	5			16
6	1	1	1			3
Grand Total	3	25	45	10	1	84

- Increase: 66%; Same: 30%; Decrease: 4%

4. Metaware challenge

- Are those scoring in the Metaware tier merely displaying “esoteric spiritual language”
- or do they also **have sufficient 4.5/Strategist** scores, showing they include that complexity?

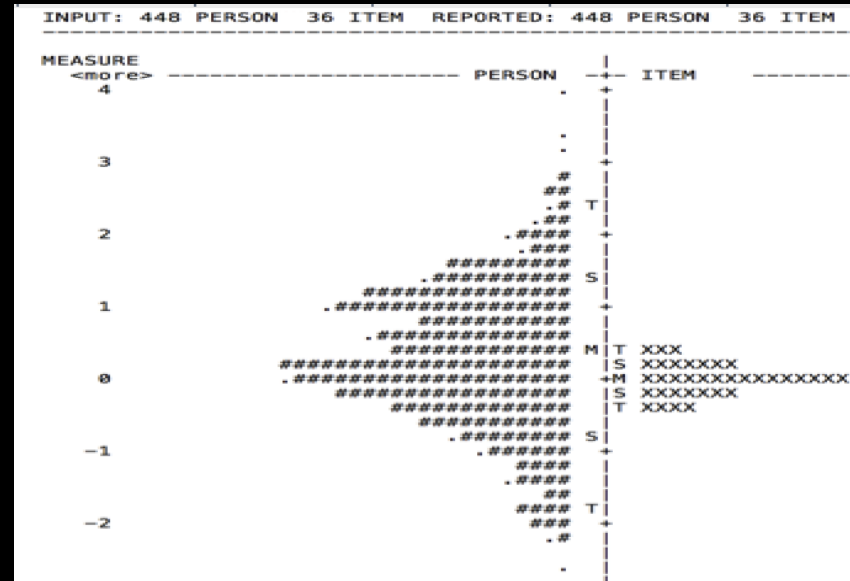
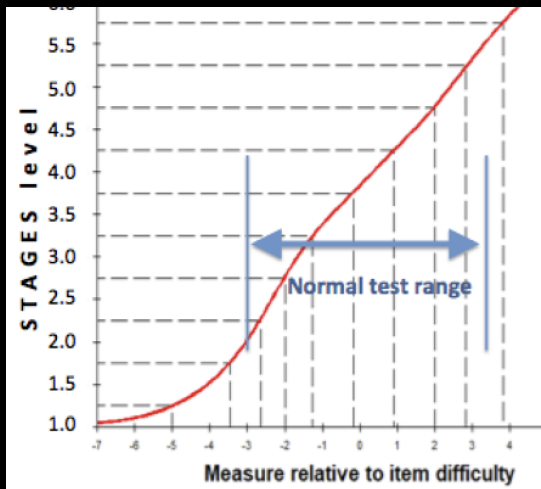
Percent of items at each level



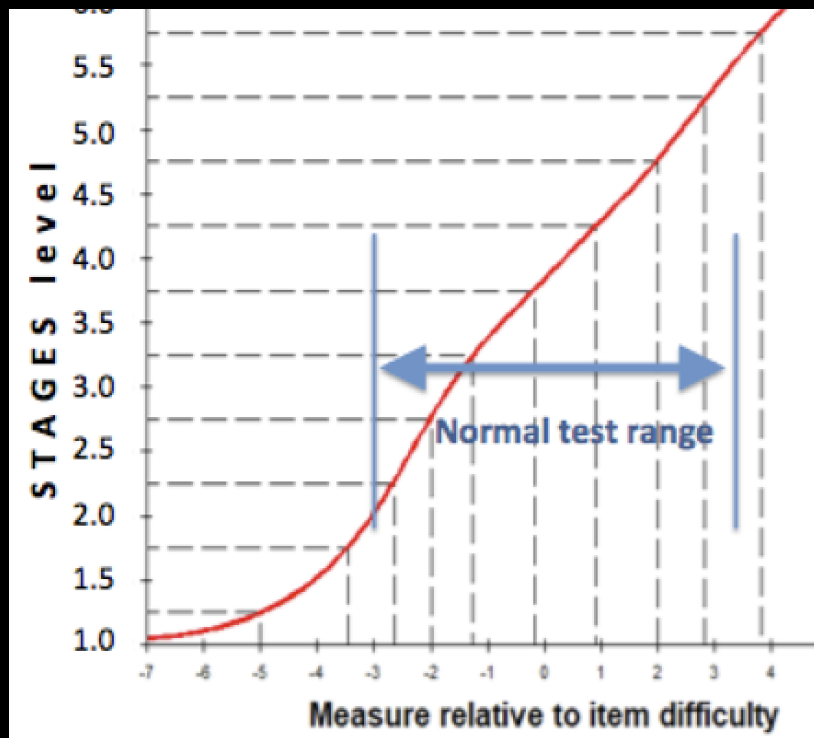
- 5.0, 5.5, 6.0 have expected values at 4.5

5. Rasch Analysis at item level

- Q: How does STAGES validate through modern Rasch (and “item response theory”) validation?
- RESULTS:
 - Test looks strong
 - Levels are evenly spaced



Rasch Analysis at item level



INPUT: 448 PERSON 36 ITEM REPORTED: 448 PERSON 36 ITEM 12

MEASURE	PERSON	ITEM
4	.	+
	.	+
	.	+
3	#	+
	##	+
	.#	T
	##	+
2	.####	+
	.###	+
	#####	+
	.#####	S
	#####	+
1	.#####	+
	#####	+
	.#####	+
	#####	M T XXX
	#####	S XXXXXXXX
0	.#####	+M XXXXXXXXXXXXXXXX
	#####	S XXXXXXXX
	#####	T XXXX
	#####	+
-1	.#####	+
	####	+
	.####	+
	##	+
	####	T
-2	###	+
	.#	+
	.	+
-3	.	+

<more> ----- PERSON +- ITEM -----

EACH "#" IN THE PERSON COLUMN IS 2 PERSON: EACH "." IS 1

<less> ----- PERSON +- ITEM -----

6. Cutoff method of item aggregation: drawbacks of “ogive” method

- Lack of empirical support for discontinuous levels
- No population studies for Bayesian estimates,
- Difficulties estimating extreme values
- Lower cutoffs that confuse development with "shadow" evidence
- Sensitivity to error
- Errors in the original ogive formula and its assumptions

New method to calculate “center of gravity” from the 36 (or fewer) stems

- “Core Score” – based on average of the top 6 scores
- Gives continuous (i.e. fractional) results
- Includes a “Bottom” (or shadow) score
- Does not confuse early development with “shadow crash” items
- Allows studying children (see below)
- Fewer ad-hoc assumptions
- Extensible

7. Specialty Inventories

- “What does development look like in domain X?”
- STAGES allows for easy creation of new stems and topic-focused inventories without re-writing the scoring manual.
- Steps for scored protocols: (1) internal consistency validation; (2) qualitative analysis of text for themes

Education	validated
Love	validated
Climate Change	validated
Dementia	validated
Hope	validated
Leadership/Business	validated
children	validated
Relationships	validated
Money	validated
Psychology	validated
Mormon (LDS) spirituality	data-gathered
Health Care for Chronic Illness	data-gathering
Alternative (incl.Chinese) Medicine	data-gathering
Resilience	data-gathering
Anomalous experiences	
Ethics	
Gender	

Specialty Inventories- Validity (Cronbachs alpha)

Table 2. Cronbach's alpha internal consistency values for specialty protocols.

Domain	<i>Cronbach's alpha</i>					
	Leadership	Love	Education	Climate Change	Psych	Children
N in study	32	20	20	32	17	53
New stems (of 36)	6	6	6	6	5	11
α All 36 stems	0.97	0.95	0.95	0.96	0.92	0.88
α General <u>prot.</u> stems only	0.96	0.94	0.94	0.95	0.90	0.85
α New stems only	0.85	0.82	0.83	0.82	0.80	0.63

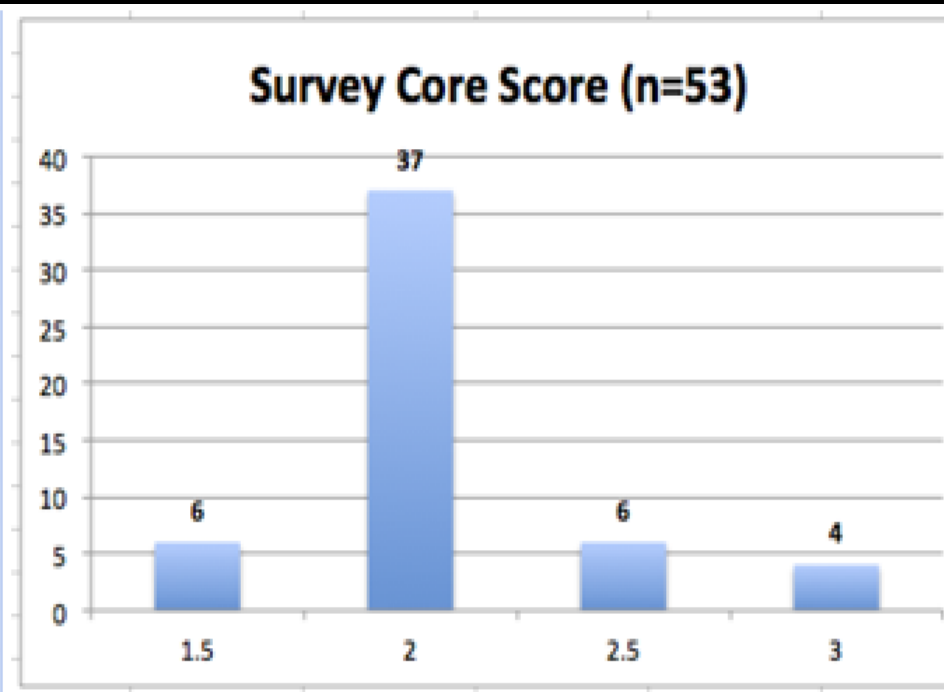
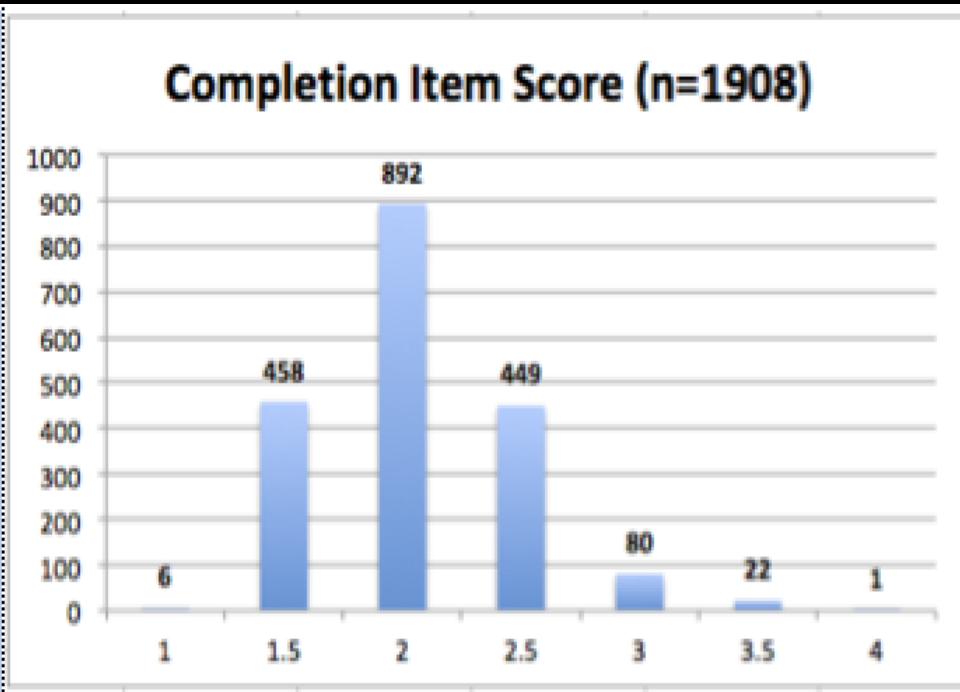
- (Excluding Childrens) The **overall** internal consistency was **excellent** (α 0.95 to 0.97).
- Internal consistency for the **new stems** as a group was **good** (α 0.80 to 0.85).

Specialty Inventories: Conclusions

- *(For the five specialty inventories excluding Children's)*
- All had excellent internal validity for both the full inventory and the subset of specialty items.

8. Children's Development

n=53, ages 4-13
(verbal completions, transcribed)



- First study of its kind...analysis in progress
- Made possible by our new Aggregation method

9. Vocabulary Analysis (preliminary)



Figure 3. Sample Word Frequencies.

10. “Face Validity” of STAGES

- Use in about a dozen PhD research studies
- Other applications and uses...

PhD research using STAGES

- Lynam, A. - **Teaching** and Mentoring in Higher Education
- Hochachka, G. - Importance of human interiority in transformative approaches to **climate change**
- Mantler, N. - **Women's** Authentic **Leadership** Development
- Churchill, J. - Indo-Tibetan **Contemplative** Essence Psychology
- Braks, A. - Transformational Executive **Coaching**
- Buckley, L. - **Hope** Examined Through a Developmental Stage Perspective
- Reynolds, E. - Next-Stage **Organizations**: A Transdisciplinary Case Study
- Attebery, J. - **Sustainability** Pedagogy for Incarcerated Adult Learners.
- Parker, J. - STAGES of **Organizational Development**.
- Hurst, R. - Ego Development and Self Initiated **Reflection**
- Aguocha, F. – **Relational Leadership** and Ego Development

Masters Theses and other applications

- Miller, J. - Importance of **Reflective Practice** in Self-Knowledge.
- John Kesler's "**Integral Polarity Practice**" (IPP)
- Marj Britt, Senior Minister Emeritus in the Unity Church, work on "the soul's journey through life [as a] **cosmic and human love story**."
- Pacific integral's GTC (**Generating Transformative Change**) cohort program (Geoff Fitch and Abigail Lynam)
- Translations to **German Language** (in process)
- Hilbert, M. - **Susceptibility to Fake News** and media manipulation

RECAP: STAGES Research

1. *STAGES vs. SCTi-MAP Replication Study*
2. Inter-Rater Studies at the item scale
3. *Longitudinal Analysis*
4. Analysis of Late Stage Patterns
5. Rasch Analysis of item and test strength
6. Investigating the validity of the Ogive Aggregation Method
7. *Specialty Inventories – custom SCTs*
8. *Studies of Children's Inventories*
9. *Vocabulary Analysis*
10. Face Validity: STAGES Applications

