

Psychometrics in a Learning Maps Environment: Item Analysis

Amy Clark

University of Kansas



The present publication was developed under grant 84.373X100001 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the author(s), and no official endorsement by the U.S. Department should be inferred.



DYNAMIC
LEARNING MAPS

Overview

- Nodes in the learning map are the key psychometric pieces measured by the assessment system
- The psychometric model relies on cognitive diagnostic methods to provide the probability of node mastery rather than a continuous scaled score
 - See Dr. Templin’s presentation for slides on the model

Items

- Items are organized into testlets
- Testlets consist of an engagement activity and 3-8 items
- Model includes testlet effect
- Constraints:
 - Sample size
 - Multiple item types
 - Writing testlets assess nodes by answer option rather than item

ITEM ANALYSIS

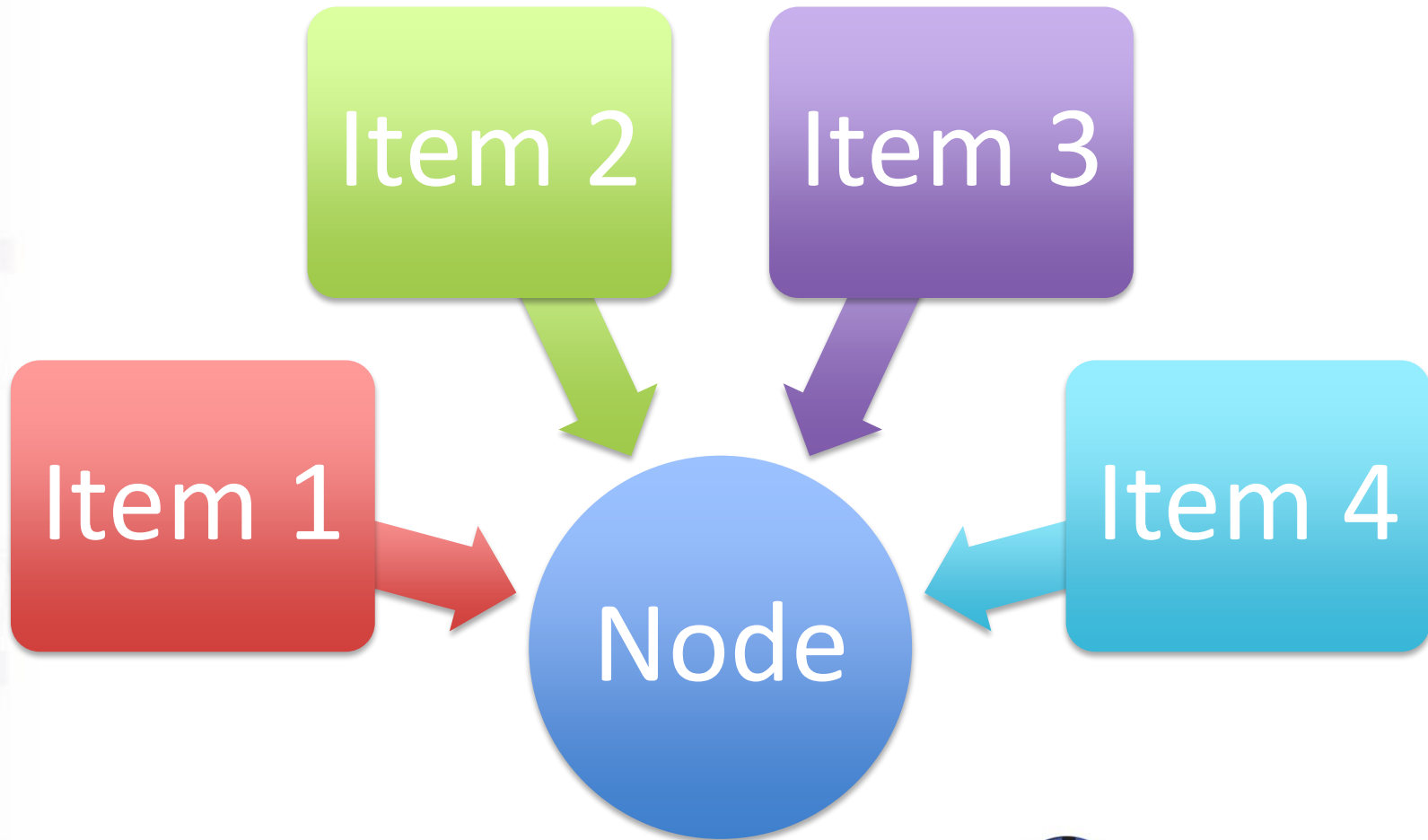


DYNAMIC
LEARNING MAPS

Item Analysis

- Item analysis is node-based, examining how items within and across nodes in the learning map perform
- Items are analyzed within the context of the node(s) they are intended to assess
- Traditional flagging criteria are modified to account for items nested within nodes

Item Analysis



Item Analysis

- Interested in similar concepts:
 - Difficulty
 - Discrimination

Traditional

- p value
- PBS correlation
- a , b , c parameters

Dynamic Assessment

- p value
- Intercept
- Main effect

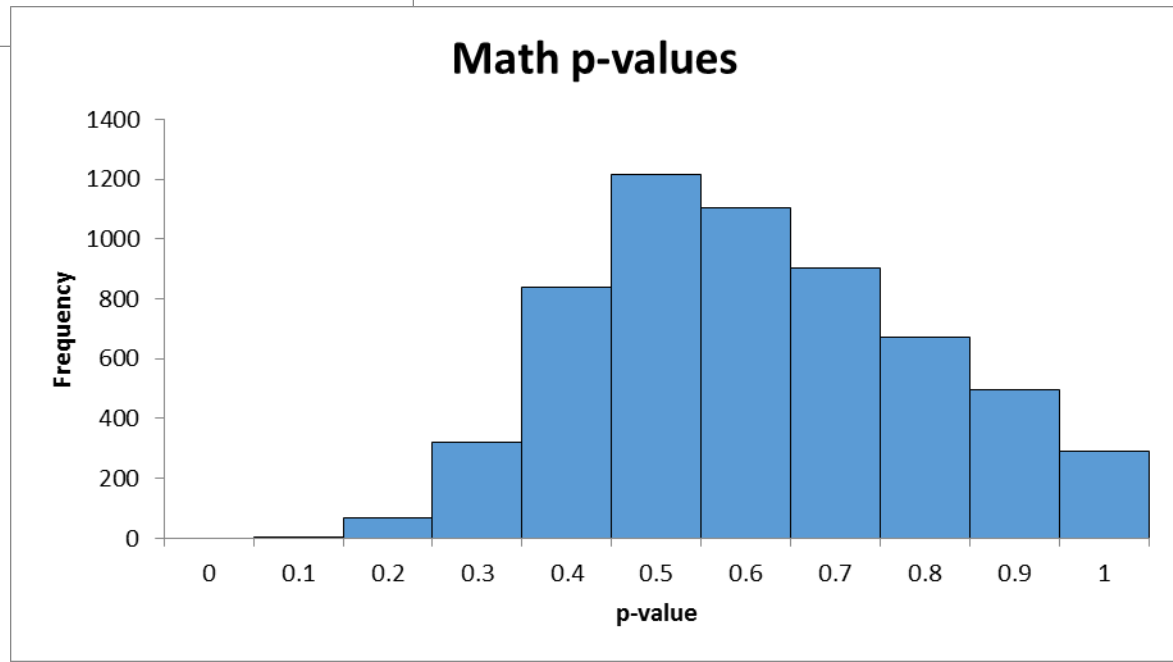
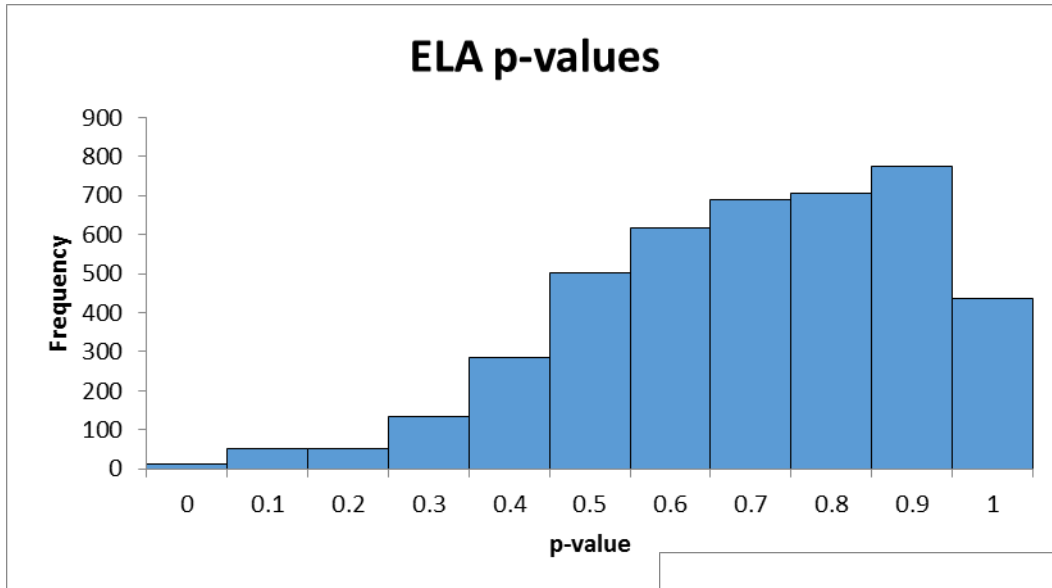
Item Analysis

- Five flagging areas:
 1. Difficulty
 2. Fungibility
 3. Non-informative items
 4. Reversals
 5. Overspecification

1. Difficulty

- Items measuring the node should be appropriately challenging
 - If too difficult, perhaps items are not measuring the skill intended
- Most DLM items are 3-option multiple-choice
- Given an historic lack of opportunity to learn we set threshold for p value flagging at 0.35

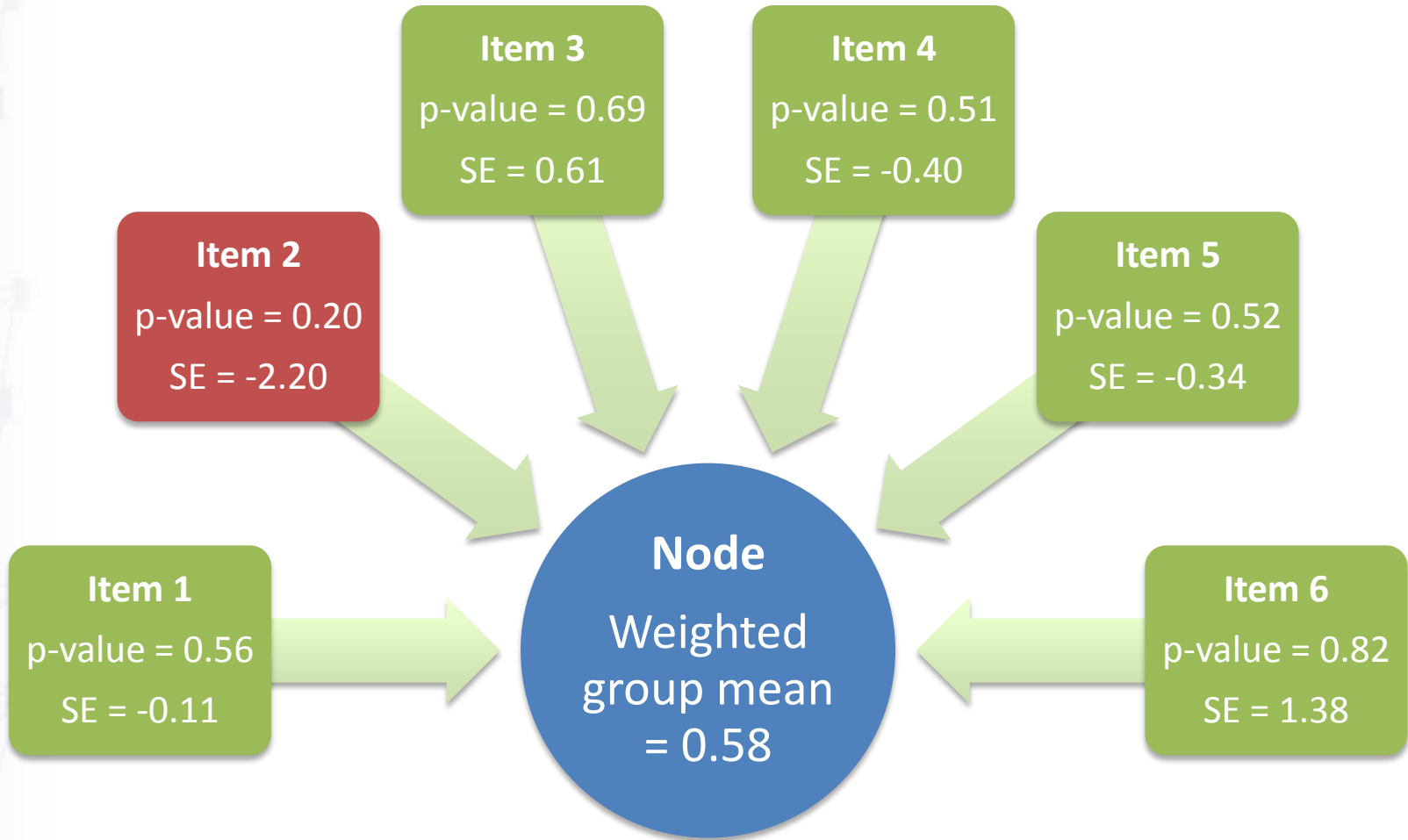
1. Difficulty



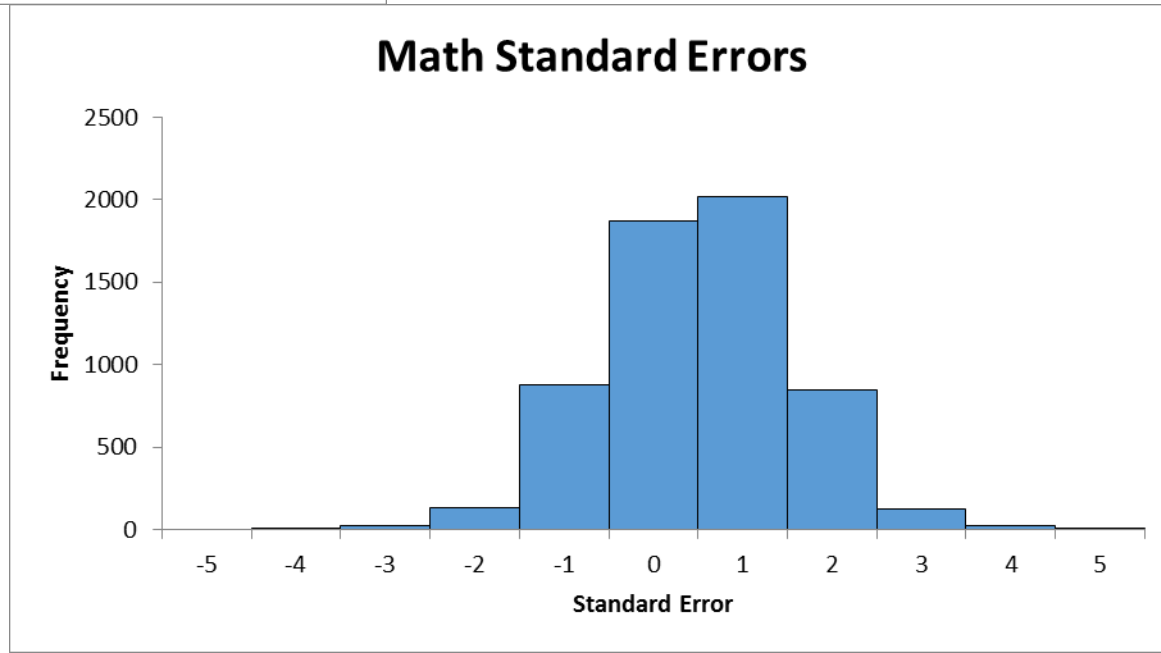
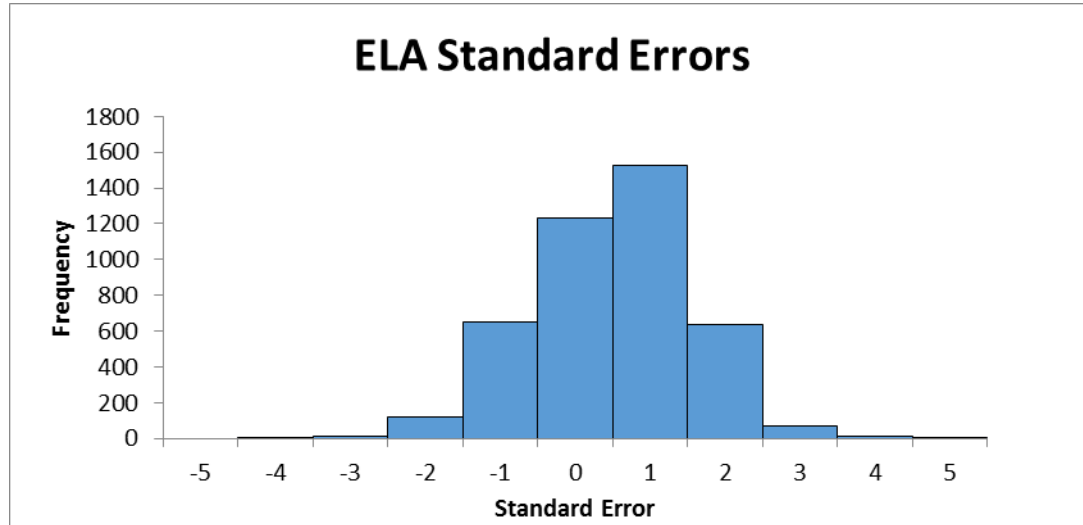
2. Fungibility

- All items measuring the node should be of approximately equal difficulty
- It should not be of consequence to the student which item is received during the assessment
- Flagging criteria: Item p-value ≥ 1.96 standard errors different than the weighted mean of all item p-values assessing the node

2. Fungibility



2. Fungibility



Spring 2014 Flags: ELA

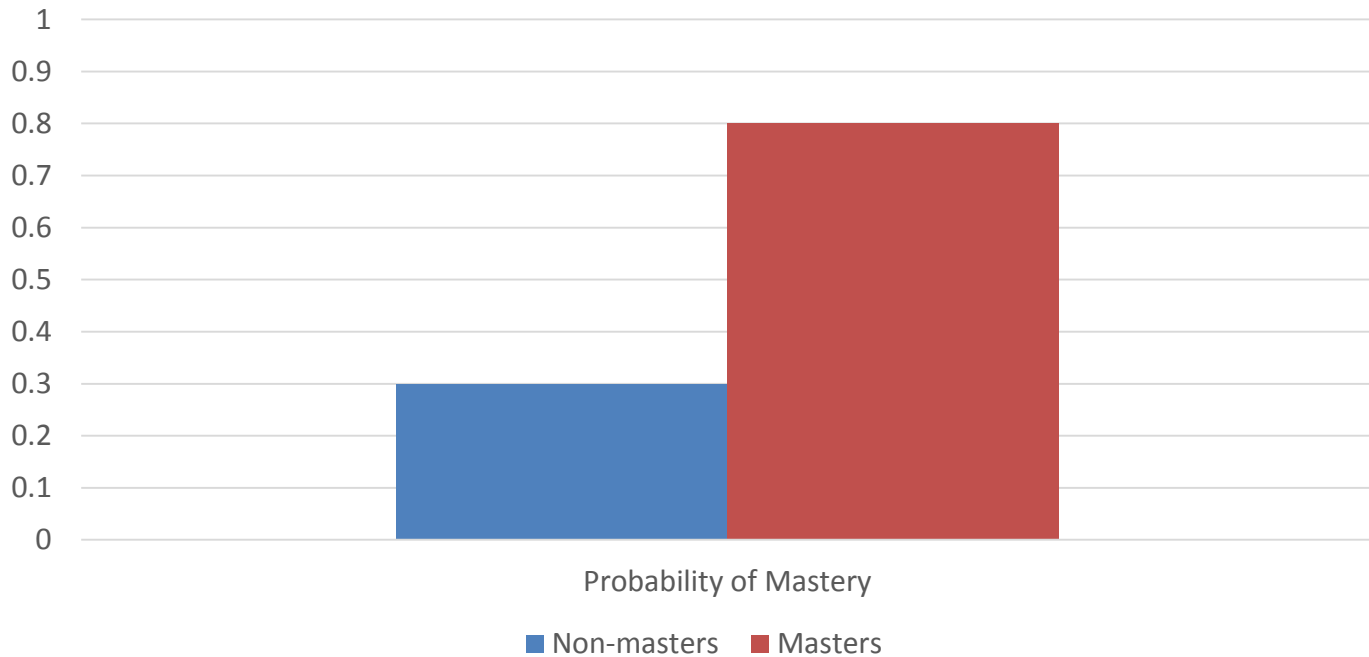
Grade	Flags	Total Items	Percent Flagged
3 rd	42	241	17.4%
4 th	16	218	7.3%
5 th	7	230	3.0%
6 th	9	216	4.2%
7 th	40	278	14.4%
8 th	25	226	11.1%
9 th – 10 th	26	233	11.2%
11 th – 12 th	13	283	4.6%
Total	178	1925	9.2%

Spring 2014 Flags: Math

Grade	Flags	Total Items	Percent Flagged
3 rd	35	308	11.4%
4 th	27	319	8.5%
5 th	50	306	16.3%
6 th	31	302	10.3%
7 th	63	424	14.9%
8 th	35	299	11.7%
9 th – 12 th	70	353	19.8%
Total	311	2311	13.5%

3. Non-Informative Items

- Items should discriminate between masters and non-masters of the node



3. Non-Informative Items

- Items are flagged as “non-informative” at the node level when:
 1. The item intercept value is greater than 1
- If the item intercept parameter is greater than 1, roughly 73% of non-masters of the node have provided a correct response

3. Non-Informative Items

- And/or items are flagged as “non-informative” at the node level when
 1. The item main effect value is less than 0.5
 2. The item main effect value is less than 0.5
- A low main effect value indicates node masters do not have a notable increase in the odds of providing a correct response

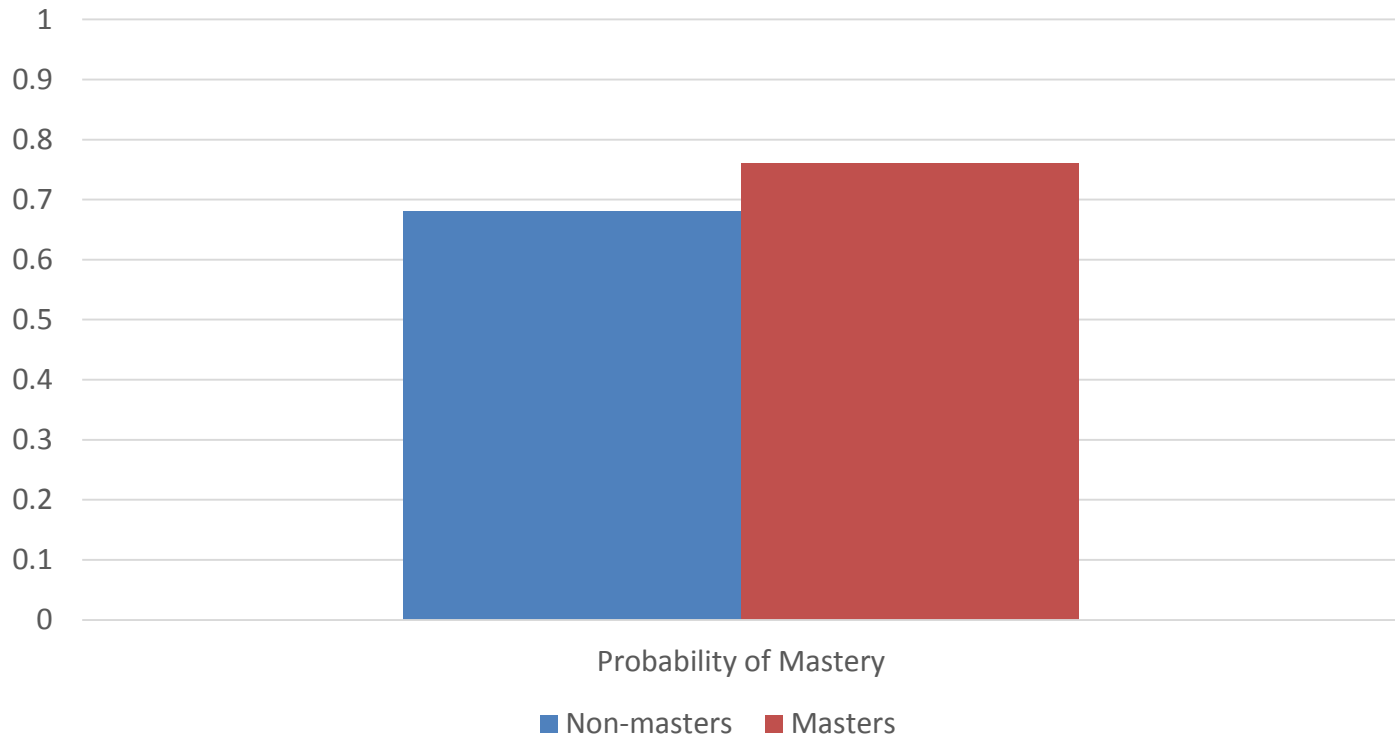
3. Non-Informative Items

4 of 12 items measuring node ELA-1141

Items	Intercept	Main Effect	Non-master		Master	
			Correct	Incorrect	Correct	Incorrect
21793	0.30	0.38	0.68	0.32	0.76	0.24
26188	-0.75	1.80	0.46	0.54	0.84	0.16
26591	0.46	264.03	0.74	0.26	1.0	0.0
26992	-0.24	24.50	0.58	0.42	1.0	0.0

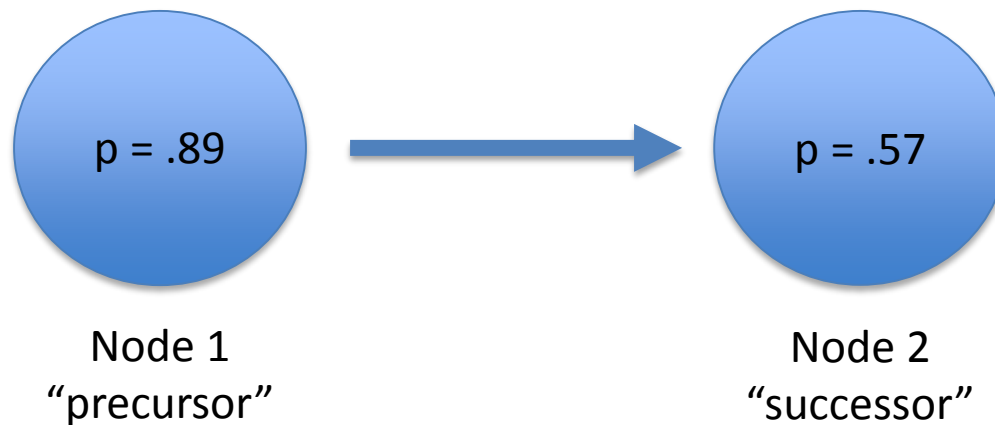
3. Non-Informative Items

Item 21793 measuring node ELA-1141



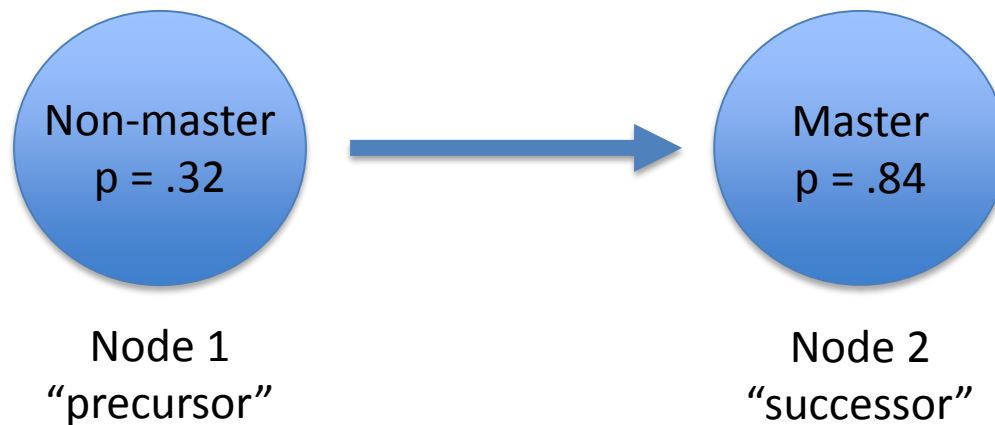
4. Node Reversals

- Students should master nodes in the order specified by the map



4. Node Reversals

- Reversal flags occur when non-masters have a high chance of being a master on subsequent nodes



4. Node Reversals

- For a node reversal to be present, the intercept value of the successor node must be greater than zero
- Could be a feature of the item(s) that is causing the reversal
- Alternately, could be an issue with map specification

4. Node Reversal

Nodes	Intercept	SD	Probability	Heidelberger p-value
ELA 1025	0.37	0.70	0.59	0.13
ELA 1110	0.04	0.18	0.51	0.40
ELA 1128	0.10	0.20	0.53	0.95
ELA 1136	3.02	1.38	0.95	0.06
ELA 1147	1.44	0.18	0.81	0.17


4. Node Reversal

Nodes	Intercept	SD	Probability	Heidelberger p-value
ELA 1025	0.37	0.70	0.59	0.13
ELA 1110	0.04	0.18	0.51	0.40
ELA 1128	0.10	0.20	0.53	0.95
ELA 1136	3.02	1.38	0.95	0.06
ELA 1147	1.44	0.18	0.81	0.17

4. Node Reversal

Nodes	Intercept	SD	Probability	Heidelberger p-value
ELA 1136	3.02	1.38	0.95	0.06
ELA 1147	1.44	0.18	0.81	0.17

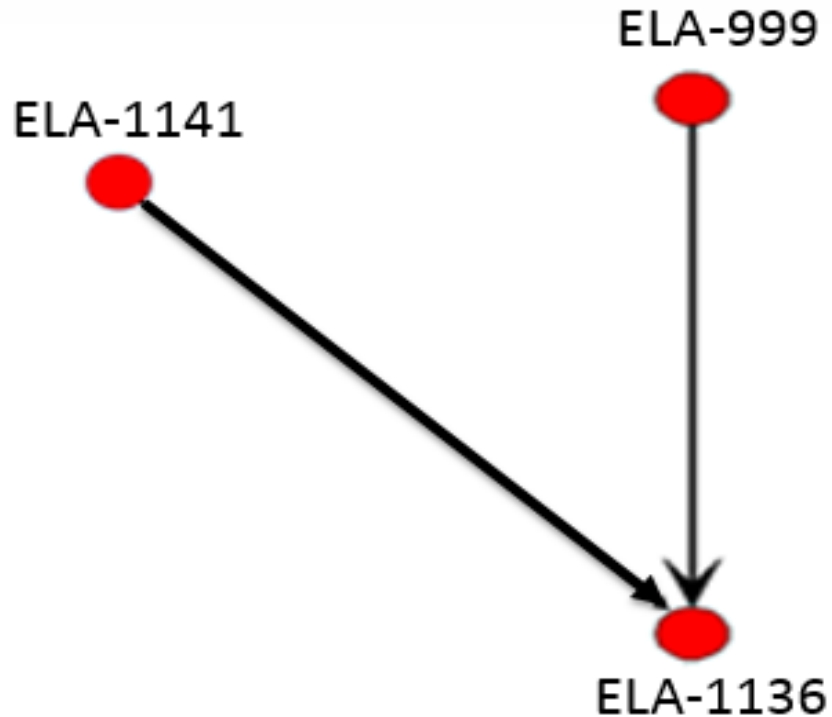
4. Node Reversal



Nodes	Intercept	SD	Probability	Heidelberger p-value
ELA 1136	3.02	1.38	0.95	0.06
ELA 1147	1.44	0.18	0.81	0.17

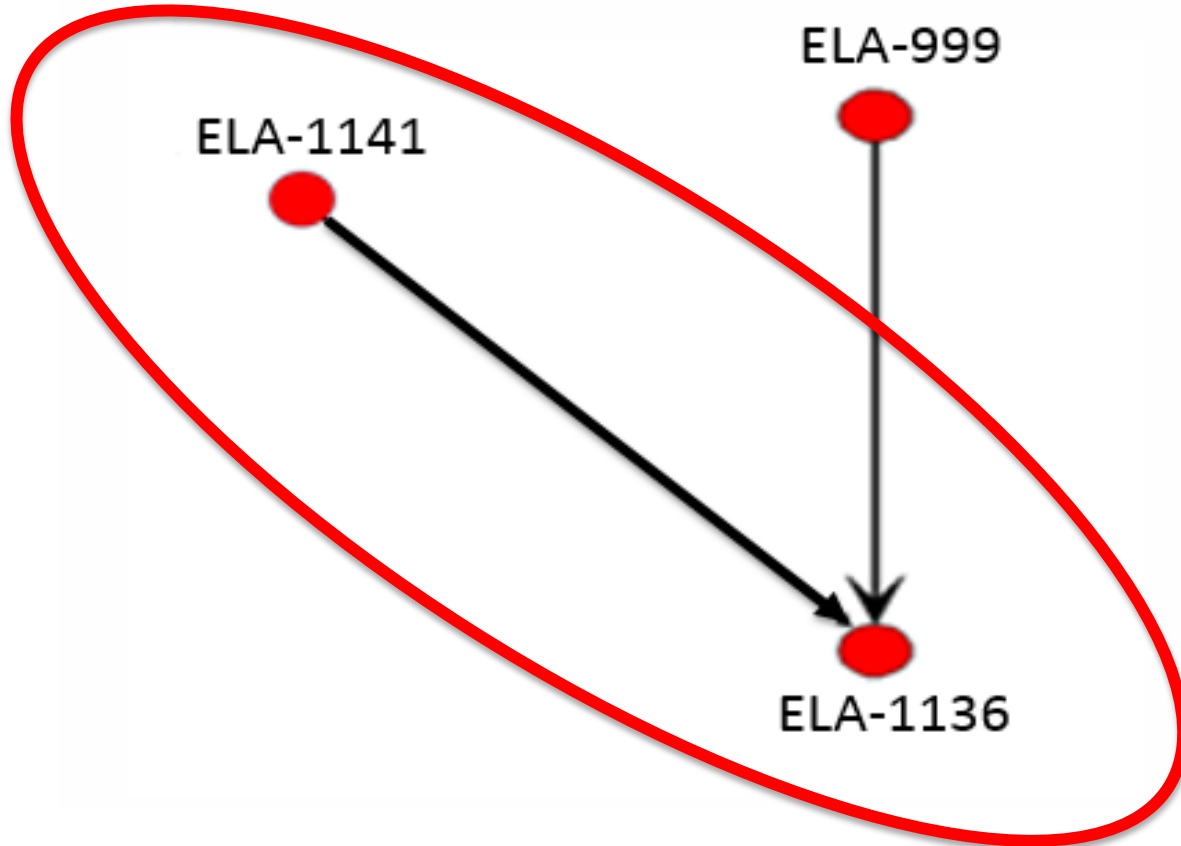
4. Node Reversal

Precursor nodes for ELA-1136



4. Node Reversal

Precursor nodes for ELA-1136

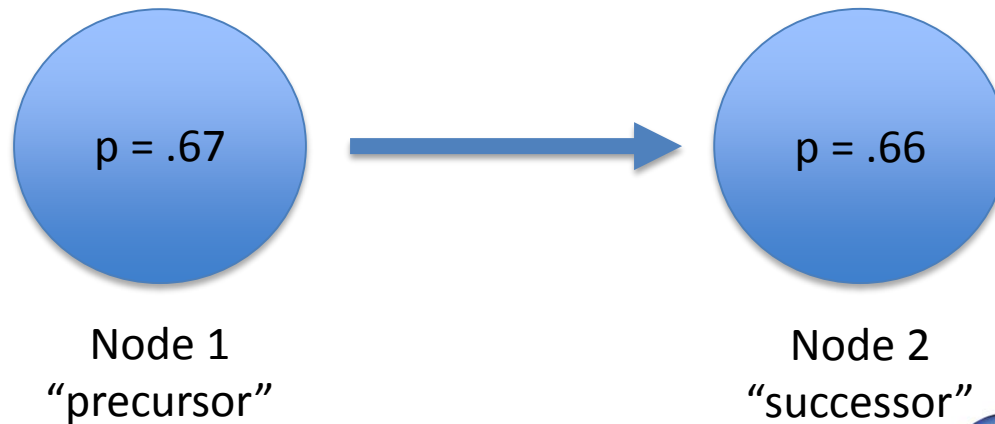


4. Node Reversal

	ELA- 1136	
ELA - 1141	Master	Non-Master
Non-Master	.96	.04
Master	.99	.01

5. Node Overspecification

- Occurs when nodes are not distinct from one another
 - Performance on successor node is not different from performance on precursor node



5. Node Overspecification

- We define an overspecified node to mean any successor node where:
 - The intercept value of the node is less than -4, and,
 - The main effect value of the node is greater than 8
- Again, could be a feature of the item causing the appearance of overspecification

CONTENT REVIEW OF FLAGGED ITEMS



DYNAMIC
LEARNING MAPS

Content Review

- Statistical information helps test developers identify items that need further review
- Statistical criteria for item selection are guidelines and cannot override content specifications
- A flag does not necessarily mean the item is “bad”

Content Review

- Item flags are reviewed by content teams
 - Each item is reviewed individually and in the context of the testlet
 - The number of items in a testlet that are flagged is also considered
- Decisions are made to:
 - Accept the item without further changes
 - Reject the item
 - Revise the item

Item Acceptance

- Item is consistent with DLM standards
- Item is aligned to the node
- Item is performing as expected given the content being assessed
 - E.g. 5-option multiple choice with p-value flag

Item Rejection

- Item is inconsistent with DLM standards
- EE and linkage level is covered by other testlets that have better performing items
- There is not a clear content-based revision to improve the item
- Revision would require complete item rewrite

Item Revision

- Technical revisions do not require changes to the content
- Examples include:
 - Item is mis-keyed
 - Item is placed at an inappropriate location within the testlet
 - Item's learning map node is incorrectly assigned

Item Revision

- Content revisions require changes to the item
- Examples include:
 - Distractors are too close to the correct response
 - Language in stem or answer choices
 - Lack of parallel construction in answer choices
 - Issues with item graphics

Spring 2014 Decisions: ELA

Grade	Flags	Accept	Revise	Reject
3rd	42	35	1	6
4th	16	6	1	9
5th	7	4	1	2
6th	9	5	1	3
7th	40	17	3	20
8th	25	3	1	21
9th – 10th	26	8	4	13
11th – 12th	13	6	3	4
Total	178	85	15	78
Percent of total		47.8%	8.4%	43.8%

Spring 2014 Decisions: Math

Grade	Flags	Accept	Revise	Reject
3rd	35	29	6	0
4th	27	21	5	1
5th	50	42	7	1
6th	31	15	11	5
7th	63	37	25	1
8th	35	16	15	4
9th – 12th	70	36	20	14
Total	311	196	89	26
Percent of total		63.0%	28.6%	8.4%

Next Steps After Review

- If accepted, item becomes operational
 - Item flagging criteria implemented following each testing window
- Revised items are subject to additional field testing in subsequent testing window

Map Validation

- Criteria used to evaluate the items can also be used to evaluate the map
 - Difficulty: node potentially assessed at wrong linkage level
 - Node reversals may indicate nodes are not in correct order
 - Overspecification may indicate nodes need to be collapsed

THANK YOU!

For more information, please contact:

akclark@ku.edu

or

Go to: www.dynamiclearningmaps.org



The present publication was developed under grant 84.373X100001 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the author(s), and no official endorsement by the U.S. Department should be inferred.



DYNAMIC
LEARNING MAPS