Multiple-choice assessments as practical diagnostic tools

Darryl Chamberlain Jr.

Department of Mathematics, University of Florida

FIU Math Ed Seminar, November 13th 2020

Outline

Introduction

Development

Evidence of Success

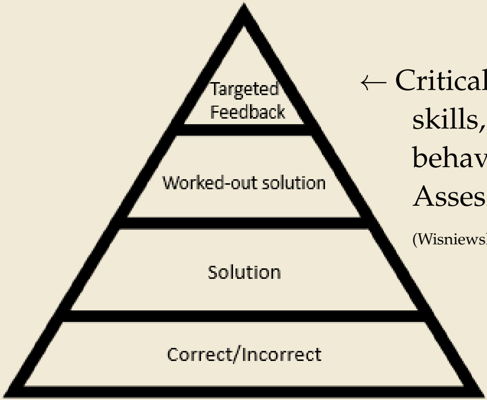
Conclusions

Competing Goals of Assessments

	as Learning	**for** Learning	**of** Learning
Learn	During	After	Before
Attempts	Single	Multiple	Single
Stakes	None	Low	High
Feedback	During	After	None

(Dann, 2014)

TYPES OF FEEDBACK

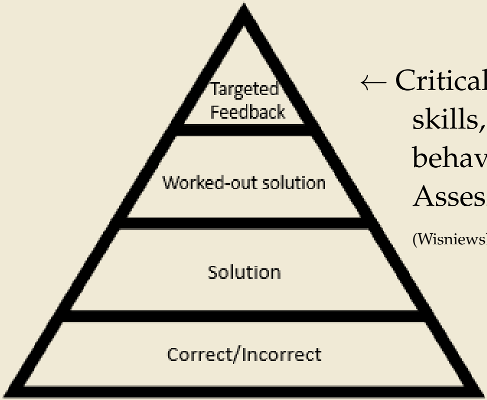


← Critical for cognitive, motor skills, motivational, and behavioral outcomes in Assessment as/for Learning

(Wisniewski et al., 2020).

Wisniewski et al. (2020); Hattie & Timperley (2007)

Types of Feedback



Wisniewski et al. (2020); Hattie & Timperley (2007)

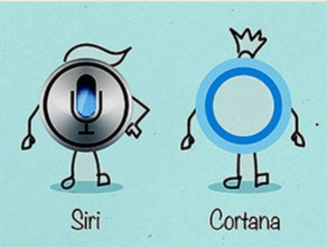
← Critical for cognitive, motor skills, motivational, and behavioral outcomes in Assessment as/for Learning (Wisniewski et al., 2020).

How do we mediate this online and/or at a large scale?

Types of Intelligence

Augmented Intelligence
Humans imbued with computer intelligence.

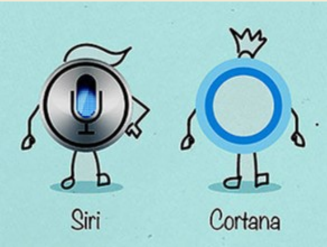
Artificial Intelligence
Computers imbued with human intelligence



Types of Intelligence

Augmented Intelligence
Humans imbued with computer intelligence.

Artificial Intelligence
Computers imbued with human intelligence



The goal is to anticipate the types of feedback an instructor would make in-the-moment and automate this feedback.

WHY MULTIPLE-CHOICE?

- Most portable and software-independent type of assessment.

Why Multiple-Choice?

- Most portable and software-independent type of assessment.
- Reduces number of responses to associate to targeted feedback.

WHY MULTIPLE-CHOICE?

- Most portable and software-independent type of assessment.
- Reduces number of responses to associate to targeted feedback.
- Foundation of multiple-choice diagnostic assessments can be utilized for free-response assessments and/or artificial intelligence.

OUTLINE

Introduction

Development

Evidence of Success

Conclusions

Anatomy of a Multiple-Choice Question

[stem] Solve the equation below.

$$\text{[problem]}\frac{-8x+9}{3}-\frac{-6x+9}{5}=\frac{-9x-9}{8}$$

A. [solution] $x = 6.805$

B. [distractor] $x = -0.332$

C. [distractor] $x = 17.341$

D. [distractor] $x = 26.341$

Types of Distractors

	Superficial	**Errors**	**Conceptions**
Frequency	Common	Uncommon	Rare
Cognitive level	None	Procedural	Conceptual
$2x^2 - x - 3 =$	$(2x-3)(x-1)$	$(2x+3)(x-1)$	Unfactorable $(x-6)(x+1)$
Explains	???	In-the-moment issues	Conceptual issues

Types of Distractors

	Superficial	**Errors**	**Conceptions**
Frequency	Common	Uncommon	Rare
Cognitive level	None	Procedural	Conceptual
$2x^2 - x - 3 =$	$(2x-3)(x-1)$	$(2x+3)(x-1)$	Unfactorable $(x-6)(x+1)$
Explains	???	In-the-moment issues	Conceptual issues

By tracking errors and misconceptions throughout an assessment/semester, students can get a better sense of what they are doing wrong and how to improve.

DEVELOPING “GOOD" DIAGNOSTIC DISTRACTORS

- Instructor’s math/teaching experience

DEVELOPING “GOOD" DIAGNOSTIC DISTRACTORS

- Instructor’s math/teaching experience
- Historical class data

DEVELOPING “GOOD" DIAGNOSTIC DISTRACTORS

- Instructor’s math/teaching experience
- Historical class data
- Historical development of concepts

Developing "Good" Diagnostic Distractors

- Instructor's math/teaching experience
- Historical class data
- Historical development of concepts
- Research results

Building Generalized Question Structures: Property

Choose the **smallest** set of Complex numbers that the number below belongs to.

$$\sqrt{\frac{-560}{5}} + \sqrt{0}i$$

A. Not a Complex Number

B. Irrational

C. * Pure Imaginary

D. Nonreal Complex

E. Rational

Property example - continued

$$\sqrt{\frac{-560}{5}} + \sqrt{0}i$$

A. Not a Complex Number

This is not a number. The only non-Complex number we know is dividing by 0 as this is not a number!

B. Irrational

These cannot be written as a fraction of Integers. Remember: Square root does not mean irrational!

C. Pure Imaginary

** This is the correct option!*

D. Nonreal Complex

This is a Complex number ($a + bi$) that is not Real (has i as part of the number).

E. Rational

These are numbers that can be written as fraction of Integers (e.g., -2/3 + 5)

Building Generalized Question Structures: Arithmetic

Solve the equation below.

$$\frac{-8x+9}{3} - \frac{-6x+9}{5} = \frac{-9x-9}{8}$$

A. * $x = 6.805$

B. $x = -0.332$

C. $x = 17.341$

D. $x = 26.341$

Arithmetic example - continued

Solve the equation below.

$$\frac{-8x+9}{3} - \frac{-6x+9}{5} = \frac{-9x-9}{8}$$

A. * $x = 6.805$; correct!

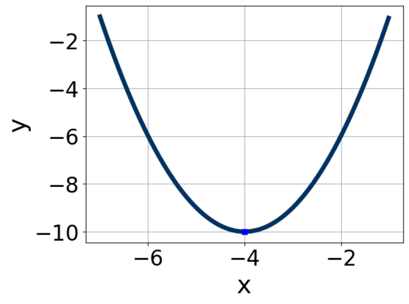
B. $x = -0.332$; corresponds to dividing the second number in the numerator by the denominator.

C. $x = 17.341$; corresponds to not distributing the negative in front of the second fraction.

D. $x = 26.341$; corresponds to dividing the coefficients in front of x by the denominator

Building Generalized Question Structures: Graphical

Choose the correct equation of the graph below. Assume $a = 1$ or $a = -1$.



A. $*f(x) = x^2 + 8x + 6$

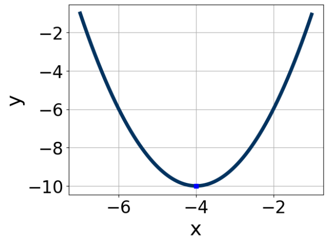
B. $f(x) = x^2 + 8x + 16$

C. $f(x) = -x^2 - 8x - 26$

D. $f(x) = x^2 + 8x + 12$

E. $f(x) = x^2 - 8x + 6$

GRAPHICAL EXAMPLE - CONTINUED



A. $*f(x) = x^2 + 8x + 6$

This is correct and uses $f(x) = a(x-h)^2 + k$.

B. $f(x) = x^2 + 8x + 16$

This uses $f(x) = a(x-h)^2 + k$ *but assumes* $k = 0$ *because it is at the bottom of the graph.*

C. $f(x) = -x^2 - 8x - 26$

This uses $f(x) = a(x-h)^2 + k$ *but assumes* $a = -1$.

D. $f(x) = x^2 + 8x + 12$

This uses a factored-form approach $f(x) = (x+6)(x+2)$ *through a visual assumption of the x-axis.*

E. $f(x) = x^2 - 8x + 6$

This uses the vertex form as $f(x) = a(x+h)^2 + k$.

BUILDING GENERALIZED QUESTION STRUCTURES: TABULAR

Given $h(x) = (f \circ g)(x)$ and **only** the information in the following table, evaluate $f(2)$, if possible.

x	$h(x)$	$g(x)$
2	1	−3
−3	4	1
−2	0	2

A. $f(2) = 4$

B. *$f(2) = 0$

C. $f(2) = 1$

D. It is not possible to evaluate $f(2)$ based only on the information in the table.

Tabular example - continued

x	$h(x)$	$g(x)$
2	1	−3
−3	4	1
−2	0	2

A. $f(2) = 4$

Corresponds to evaluating $h(g(2))$.

B. *$f(2) = 0$

Corresponds to determining the a such that $g(a) = 2$, then evaluating $h(a)$.

C. $f(2) = 1$

Corresponds to determining the a such that $h(a) = 2$, then evaluating $g(a)$.

D. It is not possible to evaluate $f(2)$ based only on the information in the table.

Corresponds to viewing functions as a single algebraic formula.

Interval Masking Method - Arithmetic example

Solve the linear equation below. Then, choose the interval that contains the solution.

$$\frac{-8x+9}{3} - \frac{-6x+9}{5} = \frac{-9x-9}{8}$$

A. $x \in [4.8, 10.8]$

B. $x \in [-2.33, 4.67]$

C. $x \in [17.34, 23.34]$

D. $x \in [23.34, 29.34]$

INTERVAL MASKING METHOD - POLYNOMIAL EXAMPLE

Construct the lowest-degree polynomial given the zeros below. Then, choose the intervals that contain the coefficients of the polynomial in the form $ax^3 + bx^2 + cx + d$.

$$\frac{3}{2}, -\frac{3}{4}, \frac{1}{2}$$

A. $a \in [10, 20], b \in [3, 7], c \in [-31, -17]$, and $d \in [1, 17]$

B. $a \in [10, 20], b \in [23, 30], c \in [-7, 1]$, and $d \in [-12, 1]$

C. $a \in [10, 20], b \in [16, 22], c \in [-17, -6]$, and $d \in [-12, 1]$

D. $a \in [10, 20], b \in [-22, -19], c \in [-17, -6]$, and $d \in [-12, 1]$

E. $a \in [10, 20], b \in [-22, -19], c \in [-17, -6]$, and $d \in [1, 17]$

Interval Masking Method - Polynomial example continued

A. $a \in [10, 20], b \in [3, 7], c \in [-31, -17],$ and $d \in [1, 17]$

$16x^3 + 4x^2 - 24x + 9$, *which corresponds to multiplying out* $(2x + 3)(4x - 3)(2x - 1)$.

B. $a \in [10, 20], b \in [23, 30], c \in [-7, 1],$ and $d \in [-12, 1]$

$16x^3 + 28x^2 - 9$, *which corresponds to multiplying out* $(2x + 3)(4x + 3)(2x - 1)$.

C. $a \in [10, 20], b \in [16, 22], c \in [-17, -6],$ and $d \in [-12, 1]$

$16x^3 + 20x^2 - 12x - 9$, *which corresponds to multiplying out* $(2x + 3)(4x - 3)(2x + 1)$.

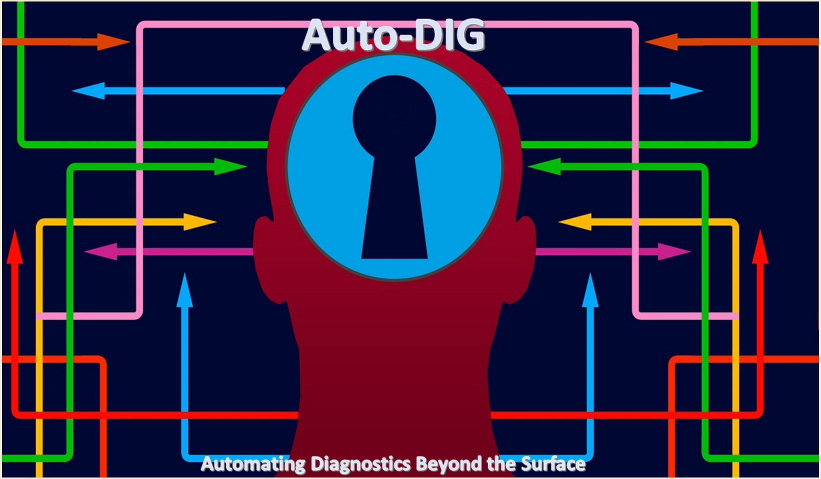
D. $a \in [10, 20], b \in [-22, -19], c \in [-17, -6],$ and $d \in [-12, 1]$

$16x^3 - 20x^2 - 12x - 9$, *which corresponds to multiplying everything correctly except the constant term.*

E. $a \in [10, 20], b \in [-22, -19], c \in [-17, -6],$ and $d \in [1, 17]$

* $16x^3 - 20x^2 - 12x + 9$, *which is the correct option.*

Graphical User Interface



Outline

Introduction

Development

Evidence of Success

Conclusions

"Solve Compound Inequality" example

Fall 2017 Q12 [$n = 76$]	**Fall 2018 Q12** [$n = 69$]	**Fall 2019 Q4** [$n = 70$]
$-1 \leq \frac{x+1}{3} \leq 3$	$7 + 7x < \frac{50x - 4}{6} \leq 8 + 8x$	$-8 + 8x < \frac{36x - 8}{4} \leq 7 + 8x$
A. $[-5, 3]$	A. $(a, b]$, where $a \in [-29, -25]$ and $b \in [-10, -5]$	A. $[a, b)$, where $a \in [-12, -7]$ and $b \in [3, 7]$
B. $[-7, 1]$	B. $[a, b)$, where $a \in [-29, -24]$ and $b \in [-10, 2]$	B. $(a, b]$, where $a \in [-10.4, -8.6]$ and $b \in [4.6, 6.4]$
C. $[-1, 7]$	*C. $(a, b]$, where $a \in [2, 8]$ and $b \in [23, 31]$	*C. $(a, b]$, where $a \in [-8.2, -4.9]$ and $b \in [6.1, 11.2]$
*D. $[-3, 5]$	D. $[a, b)$, where $a \in [4, 7]$ and $b \in [25, 29]$	D. $[a, b)$, where $a \in [-7, -2]$ and $b \in [8, 11]$
	E. There is no solution to the inequality.	E. There is no solution to the inequality.
7%, 1%, 1%, **91%**	1%, 3%, **87%**, 3%, 6%	0%, 4%, **83%**, 10%, 3%

"SOLVE LINEAR EQUATION (RATIONAL)" EXAMPLE

Fall 2017 Q9 [$n = 76$]	**Fall 2018 Q9** [$n = 69$]	**Fall 2019 Q5** [$n = 70$]
$\frac{x-2}{9} = \frac{x-4}{2}$	$\frac{-4x-6}{2} - \frac{-4x+6}{5} = \frac{3x+7}{4}$	$\frac{-8x+7}{4} - \frac{-3x+4}{5} = \frac{-3x+6}{2}$
A. $\left\{\frac{40}{7}\right\}$	A. $x \in [-1.9, -0.7]$	A. $x \in [2, 5]$
B. $\left\{-\frac{32}{11}\right\}$	*B. $x \in [-5.1, -2.1]$	B. $x \in [27, 31]$
C. $\left\{\frac{34}{7}\right\}$	C. $x \in [-11.6, -8.8]$	C. $x \in [-5, 0]$
*D. $\left\{\frac{32}{7}\right\}$	D. $x \in [-3, -1.9]$	*D. $x \in [18, 25]$
	E. No Real solutions.	E. No Real solutions.
4%, 3%, 3%, **91%**	20%, **68%**, 3%, 4%, 4%	10%, 3%, 17%, **66%**, 4%

"Factor the trinomial" example

Fall 2017 Q7 [$n = 76$]	**Fall 2018 Q6** [$n = 69$]
Factor $-24x^2 - 20x + 24$	Factor $15x^2 + 62x + 40$ as $(ax+b)(cx+d)$, with $b \le d$.
A. $-4(3x+2)(2x-3)$	A. $a \in [-4.1, -2.6], b \in [-15, -3], c \in [-7, -3]$, and $d \in [-8, -2]$
B. $4(3x+2)(2x-3)$	B. $a \in [0.7, 1.6], b \in [-3, 7], c \in [13, 17]$, and $d \in [8, 13]$
*C. $-4(3x-2)(2x+3)$	*C. $a \in [4.3, 7.8], b \in [-3, 7], c \in [-3, 6]$, and $d \in [8, 13]$
D. $4(x+1)(x-2)$	D. $a \in [0.7, 1.6], b \in [9, 13], c \in [13, 17]$, and $d \in [-1, 9]$
	E. $a \in [-4.1, -2.6], b \in [9, 13], c \in [-7, -3]$, and $d \in [-1, 9]$
13%, 1%, **86%**, 0%	0%, 1%, **93%**, 6%, 0%

Quantitative Evidence: Items with Effective Distractors

Effective distractors are those that are chosen at least 5% of the time (Hingorjo & Jaleel, 2012).

	Fall 2017				Fall 2018				Fall 2019			
	Ver A	Ver B	Ver C	AVG	Ver A	Ver B	Ver C	AVG	Ver A	Ver B	Ver C	AVG
Items w/ 0 EDs	12%	20%	32%	21%	17%	17%	13%	15%	9%	5%	0%	5%
Items w/ 1 EDs	32%	24%	32%	29%	33%	29%	25%	29%	27%	41%	41%	36%
Items w/ 2 EDs	28%	32%	28%	29%	17%	33%	25%	25%	45%	32%	23%	33%
Items w/ 3 EDs	28%	24%	8%	20%	17%	8%	25%	17%	18%	14%	27%	20%
Items w/ 4 EDs	NA	NA	NA	NA	8%	4%	4%	6%	0%	9%	9%	6%

QUANTITATIVE EVIDENCE: PERCENT OF DISTRACTORS

Semester	Fall 2017	Fall 2018/2019
# of Questions	25	24
# of Options	4	5
Total Distractors	75	96

	Fall 2017				**Fall 2018**				**Fall 2019**			
	Ver A	Ver B	Ver C	AVG	Ver A	Ver B	Ver C	AVG	Ver A	Ver B	Ver C	AVG
Ds chosen 0%	7%	17%	21%	15%	21%	22%	19%	20%	19%	27%	15%	20%
Ds chosen 0%–5%	36%	29%	41%	36%	38%	43%	38%	39%	38%	27%	34%	33%
Ds chosen 5%+	56%	53%	37%	49%	42%	34%	41%	39%	43%	45%	50%	46%

Quantitative Evidence: Distribution Statistics

“Ideal" distribution statistics:

- Center: 0.7 [5-option] or 0.74 [4-option] (Lord, 1952)
- Standard Deviation: 0.100 [5-option] or 0.087 [4-option]
- Skew and Kurtosis: 0

Negative kurtosis suggests more data to appear in the tails of the distribution (a uniform distribution has a kurtosis of -1.2) and a positive kurtosis suggests little data to appear in the tails (a logistic distribution has a kurtosis of 1.2) (Ho & Yu, 2015).

	Fall 2017				Fall 2018				Fall 2019			
	Ver A	Ver B	Ver C	AVG	Ver A	Ver B	Ver C	AVG	Ver A	Ver B	Ver C	AVG
Mean	0.76	0.77	0.81	0.78	0.76	0.83	0.74	0.77	0.77	0.77	0.76	0.77
Median	0.82	0.80	0.89	0.83	0.78	0.83	0.75	0.78	0.79	0.80	0.78	0.79
Std Dev	0.168	0.181	0.170	0.174	0.150	0.149	0.158	0.155	0.121	0.136	0.122	0.127
Skewness	-1.64	-1.15	-1.20	-1.33	-0.39	-1.63	-0.56	-0.86	-0.71	-0.96	-0.71	-0.79
Kurtosis	3.46	0.97	0.62	1.68	-1.15	4.05	0.43	1.11	0.05	0.46	-0.47	0.02

QUANTITATIVE EVIDENCE: ITEM CORRELATIONS

Point-Biserial Correlation (PBC) is an item-level correlation between marking an item correct/incorrect and a student's overall score.

Positive PBC corresponds to high-achieving students marking the question correctly while low-achieving students mark the same question incorrectly, while negative PBC corresponds to the inverse relation (Varma, 2006).

	Fall 2017				Fall 2018				Fall 2019			
	Ver A	Ver B	Ver C	AVG	Ver A	Ver B	Ver C	AVG	Ver A	Ver B	Ver C	AVG
PBCs ≤ 0	0%	0%	4%	1%	0%	4%	0%	1%	0%	0%	0%	0%
PBCs 0–0.15	8%	8%	12%	9%	0%	25%	4%	10%	5%	14%	0%	6%
PBCs 0.15–0.25	4%	13%	16%	11%	17%	8%	17%	14%	0%	14%	18%	11%
PBCs ≥ 0.25	88%	83%	72%	81%	83%	63%	79%	75%	95%	73%	82%	83%

Quantitative Evidence: Test Correlations

Kuder–Richardson Formula 20 (KR-20) reliability coefficient is used to estimate the internal reliability of an assessment (Salvucci et al., 1997).

A high KR-20 suggests the assessment would highly correlate between alternative forms

	Fall 2017				Fall 2018				Fall 2019			
	Ver A	Ver B	Ver C	AVG	Ver A	Ver B	Ver C	AVG	Ver A	Ver B	Ver C	AVG
KR-20	0.799	0.792	0.599	0.730	0.736	0.614	0.735	0.695	0.727	0.654	0.710	0.697

Overall Quantitative Evidence

Measure	Rating	Description		2017 AVG	2018 AVG	2019 AVG
Effective Distractors *Distractors chosen at 5%+.*	Poor	Items w/ 0 EDs		21%	15%	5%
	Okay	Items w/ 1 EDs		29%	29%	36%
	Good	Items w/ 2+ EDs		49%	47%	59%
Distractor Chosen *Overall percentage chosen.*	Poor	Distractors chosen 0%		15%	20%	20%
	Okay	Distractors chosen 0%–5%		36%	39%	33%
	Good	Distractors chosen 5%+		49%	39%	46%
Item Difficulty *Proportion of students answering each item correctly.*	Poor	Statistics suggest non-normal distribution.	Mean	0.78		
			Med	0.83		
			StDev	0.174		
			Skew	-1.33		
			Kurt	1.68		
Mean and Median describe the center of the distribution. Ideal: 0.7 or 0.74. *Standard Deviation describes the spread of the distribution. Ideal: 0.100 or 0.087.*	Okay	Statistics suggest somewhat normal distribution near "good" center and standard deviation.	Mean		0.77	
			Med		0.78	
			StDev		0.155	
			Skew		-0.86	
			Kurt		1.11	
Skew describes the symmetry of distribution. Ideal: 0. *Kurtosis describes the shape of the tail to the peak of a distribution. Ideal: 0.*	Good	Statistics suggest normal distribution near "good" center and low standard deviation.	Mean			0.77
			Med			0.79
			StDev			0.127
			Skew			-0.79
			Kurt			0.02
Point Biserial Correlation *Correlation between score on item and total score.*	Poorest	Below 0		1%	1%	0%
	Poor	Between 0 and 0.15		9%	10%	6%
	Okay	Between 0.15 and 0.25		11%	14%	11%
	Good	0.25 and above		81%	75%	83%
KR-20 *Correlation of alternative forms.*	Poor	0.5 or below; Between 0.9 and 1.0				
	Good	Between 0.5 and 0.8		0.730	0.695	0.697
	Ideal	0.80 - 0.9				

Outline

Introduction

Development

Evidence of Success

Conclusions

BENEFITS

- Create numerous, high-quality assessments that target **conceptions** and can be used to diagnosis/address errors.
- Frees up instructor time for more free-response time.
- Can scale up best practices in math ed.

Current Obstacles

- Developed only for College Algebra topics.
- Needs more integration of math ed research results.
- GUI currently written for Linux only.
- Students need to be taught how to answer in intervals.

Other Uses

- Open-Source Homework System

Other Uses

- Open-Source Homework System
- Intelligent Tutoring System (Stefanutti et al., 2020)

OTHER USES

- Open-Source Homework System
- Intelligent Tutoring System (Stefanutti et al., 2020)
- Large-Scale Mastery-Based Grading

OTHER USES

- Open-Source Homework System
- Intelligent Tutoring System (Stefanutti et al., 2020)
- Large-Scale Mastery-Based Grading
- Automated Assessment Reports

Large-Scale Mastery-Based Grading

10 quizzes, 2 topics per quiz, 10 questions per topic.

Q1: 20 questions per version

Q2: 40 questions per version

Q3: 60 questions per version

Q4: 80 questions per version

Q5: 120 questions per version

Q6-10: 160 questions per version

With 3 versions per exam, I end up generating...

Large-Scale Mastery-Based Grading

10 quizzes, 2 topics per quiz, 10 questions per topic.

Q1: 20 questions per version

Q2: 40 questions per version

Q3: 60 questions per version

Q4: 80 questions per version

Q5: 120 questions per version

Q6-10: 160 questions per version

With 3 versions per exam, I end up generating... 3360 Total Questions!

Large-Scale Mastery-Based Grading

10 quizzes, 2 topics per quiz, 10 questions per topic.

Q1: 20 questions per version

Q2: 40 questions per version

Q3: 60 questions per version

Q4: 80 questions per version

Q5: 120 questions per version

Q6-10: 160 questions per version

With 3 versions per exam, I end up generating... 3360 Total Questions!

Each question requires 15-30 “different" versions.

Large-Scale Mastery-Based Grading

10 quizzes, 2 topics per quiz, 10 questions per topic.

Q1: 20 questions per version

Q2: 40 questions per version

Q3: 60 questions per version

Q4: 80 questions per version

Q5: 120 questions per version

Q6-10: 160 questions per version

With 3 versions per exam, I end up generating... 3360 Total Questions!

Each question requires 15-30 “different" versions.

With ∼300 students, students would need targeted feedback to about 60,000 questions.

Automated Assessment Reports

Objective	Sub-Objective	Score	Possible Issue (if applicable)
Construct linear functions	...using a slope and point.	1/1	
	...using two points.	1/1	
	...using a parallel/perpendicular line and point.	1/1	
Convert between linear forms.	Standard to Slope-Intercept.	1/1	
	Slope-Intercept to Standard.	0/1	Standard form requires integer coefficients.
	Point-Slope to Standard.	0/1	Standard form requires integer coefficients.
Convert between linear representations.	Graph to Equation.	1/1	
	Equation to Graph.	1/1	
Solve linear equations	...with integer coefficients.	1/1	
	...with rational coefficients.	0/1	Did not distribute the negative in front of the second fraction.

RESOURCES

Contact: dchamberlain31@ufl.edu

Github:
https://github.com/Darryl-Chamberlain-Jr/Auto-DIG

Open-Access HW System:
https://xronos.clas.ufl.edu/ufmac1105

Mastery-Based Course:
https://ufl.instructure.com/courses/408087

Papers:
https://people.clas.ufl.edu/dchamberlain31/current-projects/

References I

Dann, R. (2014). Assessment as learning: blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 149-166. doi: 10.1080/0969594X.2014.898128

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *88*(1), 81–112.

Hingorjo, M., & Jaleel, F. (2012). Analysis of one-best mcqs: The difficulty index, discrimination index and distractor efficiency. *JPMA-Journal of the Pakistan Medical Association*, *62*(2), 142–147.

References II

Ho, A. D., & Yu, C. C. (2015). Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement*, *75*(3), 365–388.

Lord, F. M. (1952). The relation of the reliability of multiple-choice tests to the distribution of item difficulties. *Psychometrika*, *17*(2), 181–194.

Salvucci, S., Walter, E., Conley, V., Fink, S., & Saba, M. (1997). *Measurement error studies at the national center for education statistics*. U.S. Department of Education Office of Educational Research and Improvement.

References III

Stefanutti, L., de Chiusole, D., Gondan, M., & Maurer, A. (2020). Modeling misconceptions in knowledge space theory. *Journal of Mathematical Psychology*, *99*, 102435. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0022249620300791` doi: https://doi.org/10.1016/j.jmp.2020.102435

Varma, S. (2006). *Preliminary item statistics using point-biseral correlation and p-values*. Morgan Hill CA: Educational Data Systems Inc.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2019.03087