www.InnovationsInTesting.org

Investigating Response Time in Improving Test Design and Analysis

Russell Smith,
Senior Psychometrician,
Alpine Testing Solutions
James B. Olsen,
Senior Psychometrician,
Alpine Testing Solutions

ATP Innovations in Testing 2007
Palm Springs, California

Selected Research on Test/Item Response Time Analysis

- Giraud, Gerald and Smith, Russell W. (2005). *The effect of response time patterns on ability estimation in high stakes computer adaptive testing*.
- Hornke, Lutz F. (2000). Item response time in computerized adaptive testing. *Psicologica* (2000), 21, 175-189.
- Meyer, J. Patrick and Wise, Steven L. (2005). *Including item response time in a distractor analysis via multivariate kernel smoothing*. Paper presented at the Annual meeting of the National Council on Measurement in Education, San Francisco, CA, March, 2005.
- Schnipke, Deborah L. and Scrams, David J. (1999). *Exploring Issues of Test Taker Behavior: Insights Gained from Response Time Analyses*. Law School Admissions Council Computerized Testing Report 98-09, March, 1999.
- Schnipke, Deborah L. and Scrams, David J. (1999). *Representing Response-Time Information in Item Banks*. Law School Admissions Council, Computerized Testing Report 97-09. May 1999.
- Smith, Russell W. (2000, April). *An exploratory analysis of item parameters and characteristics that influence item response time*. Paper presented at the Annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Thissen, D. (1983). Timed testing: An approach using item response theory. In D. J. Weiss (Ed.), *New Horizons in Testing: Latent trait theory and computerized adaptive testing*. (pp. 179-203). New York: Academic Press

ATP Innovations in Testing 2007 Palm Springs, California

Selected Research on Test/Item Response Time Analysis

- Schnipke, D. L. (1995). *Assessing speededness in computer-based tests using item response theory*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April, 1998.
- Scrams, D. J. and Schnipke, D. L. (1997). *Making use of response times in standardized tests: Are accuracy and speed measuring the same thing?* Paper presented at the Annual meeting of the American Educational Research Association, Chicago, March, 1997.
- van der Linden, Wim J. , Scrams, David J. and Schnipke, Deborah L. (1998).*Using Response-Time Constraints in Item Selection to Control for Differential Speededness in Computerized Adaptive Testing*. Research Report 98-06, University of Twente. Department of Educational Measurement and Data Analysis.
- van der Linden, Wim J. (2006b). *Normal models for response times on Test Items*. Law School Admissions Council, Computerized Adaptive Testing Report 04-08, May, 2006.
- van der Linden, Wim J. (2006a). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, 31(2), 181-204.
- Wang, T. and Hanson, B. A. (2001). *Development and Calibration of an Item Response Model that Incorporated Response Time*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, April, 2001.

- Wang, T. (2006). *A model for the joint distribution of item response and response time using a One-Parameter Weibull distribution*. Center for Advanced Studies in Measurement and Assessment, CASMA Research Report No. 20. Lindquist Center, Iowa City, IA: University of Iowa.
- Wise, Steven L. and Kingsbury, G. Gage. (2005). An investigation of item response time distributions as indicators of comprimised NCLEX item pools. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL: April, 2005.
- Wise, Steven L. and Kong, Xiaojing (2005). Response time effort: A new Measure of examinee motivation in computer-based tests. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, April, 2005. In press Applied Measurement in Education.
- Zeniski, April L. and Baldwin, Peter (2006). *Using response time data in test development and validation: Research with beginning computer users*. Paper presented at the Annual meeting of the National Council on Measurement in Education, San Francisco, April, 2006.

ATP Innovations in Testing 2007
Palm Springs, California

Item/Test Response Time Applications (Schnipke and Scrams, 2002).

- Model item/test response times within the framework of item response theory (Roskam, 1997; Thissen, 1983; van Breukelen, 1989; Verhelest, Verstraalen and Jansen, 1997).
- Model item/test response times independently of the responses to items Maris, 1993; Scheiblechner. 1979; Schnipke and Scrams, 1997; van der Linden, Scrams and Schnipke, 1997; and van der Linden and Krimpen-Stoop, 2003.
- This research study uses the second response time analysis approach.

ATP Innovations in Testing 2007 Palm Springs, California

Item/Test Response Time Applications (Schnipke and Scrams, 2002 and others).

- Selecting Scoring Models
- Speed-Accuracy Tradeoff Relationships
- Examinee Strategy Use
- Test Speededness
- Test Pacing
- Predicting Finishing Times/Setting Time Limits
- Subgroup Differences
- Test Bank and Item Security (Wise and Kingsbury, 200x

ATP Innovations in Testing 2007
Palm Springs, California

Potential Scoring Models

Continuous Response Model (Samejima, 1973, 1974, 1983)

- Linear Exponential Model (Scheiblechner, 1975, 1985)
- Rule Space Model (Tatsuoka and Tatsuoka, 1980)
- Timed Response Model (Thissen, 1983)

ATP Innovations in Testing 2007
Palm Springs, California

Comparisons of Models for Speeded Tests

- Extended Rasch Models for Simple Speeded Tasks
  - Weibull Model for Momentary Ability (Verhalest, Verstralen and Jansen (1997)
  - Weibull Model for response accuracy (Roskam, 1997)
- Models for Complex Cognitive Tasks
  - Rule Space Model with Weibull distribution
  - Timed Test Model with lognormal distribution
  - Comparisons of Different Models by Schnipke and Scrams, 1999 and van der Linden 2006 a, 2006b.
    - Recommended ordering of models: lognormal, Weibull, gamma, Gaussian
  - Lognormal model recommended by Thissen, Schnipke and Scrams, van der Linden and compatible with Samejima continuous response models.
  - This research used the lognormal and normal models.

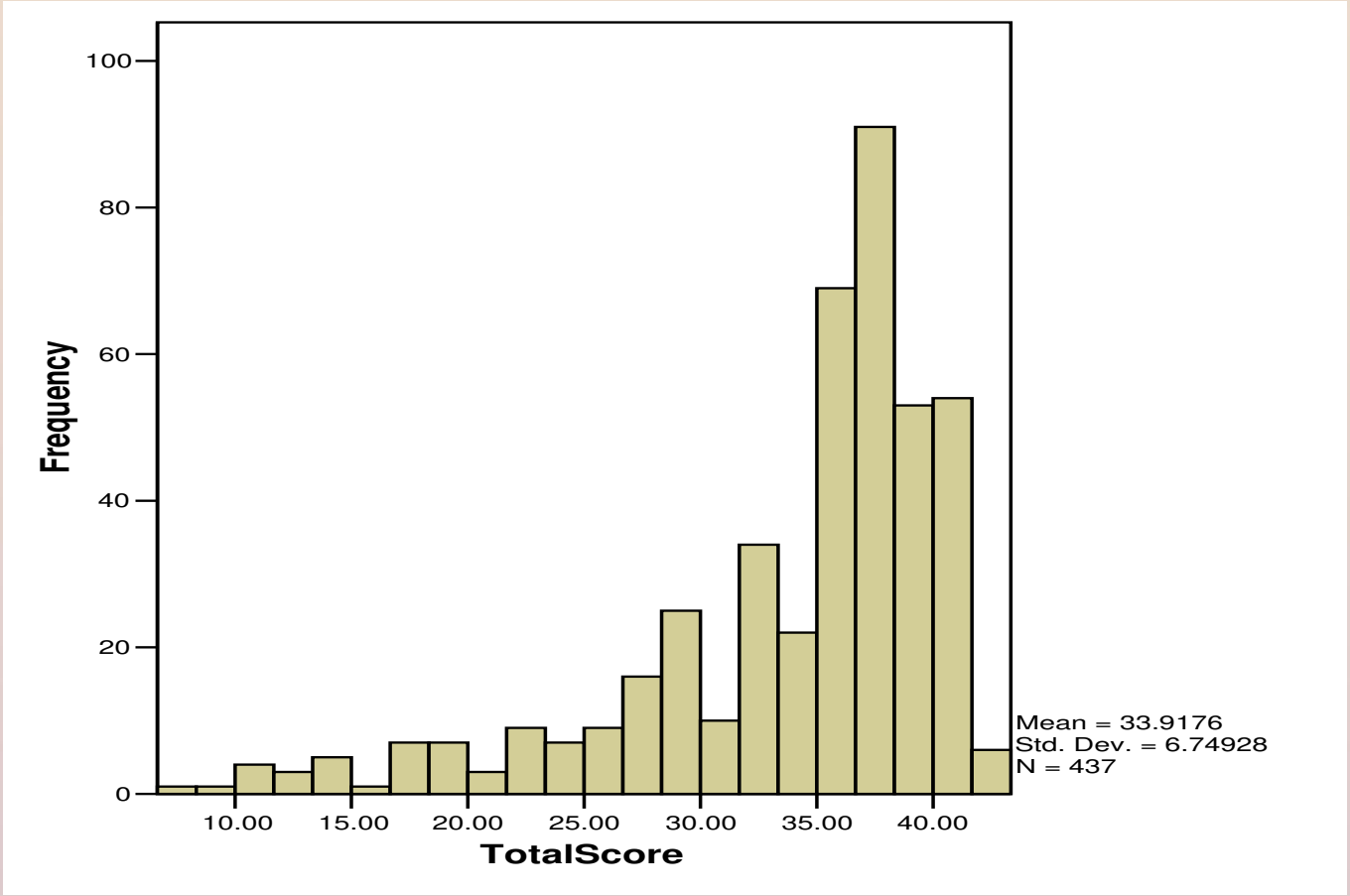
ATP Innovations in Testing 2007
Palm Springs, California

Empirical Example 2: Office Email Performance Test

TotalScore		Statistic
N Persons		437
N Items		42
Mean		33.918
Average Percent Correct		80.76%
Std. Error of Mean		0.323
Median		36
Mode		39
Std. Deviation		6.749
Variance		45.553
Alpha		.894
Standard Error Measurement		3.024
Skewness		-1.592
Std. Error of Skewness		0.117
Kurtosis		2.310
Std. Error of Kurtosis		0.233
Minimum		7
Maximum		42
Percentiles	25	32
	50	36
	75	39

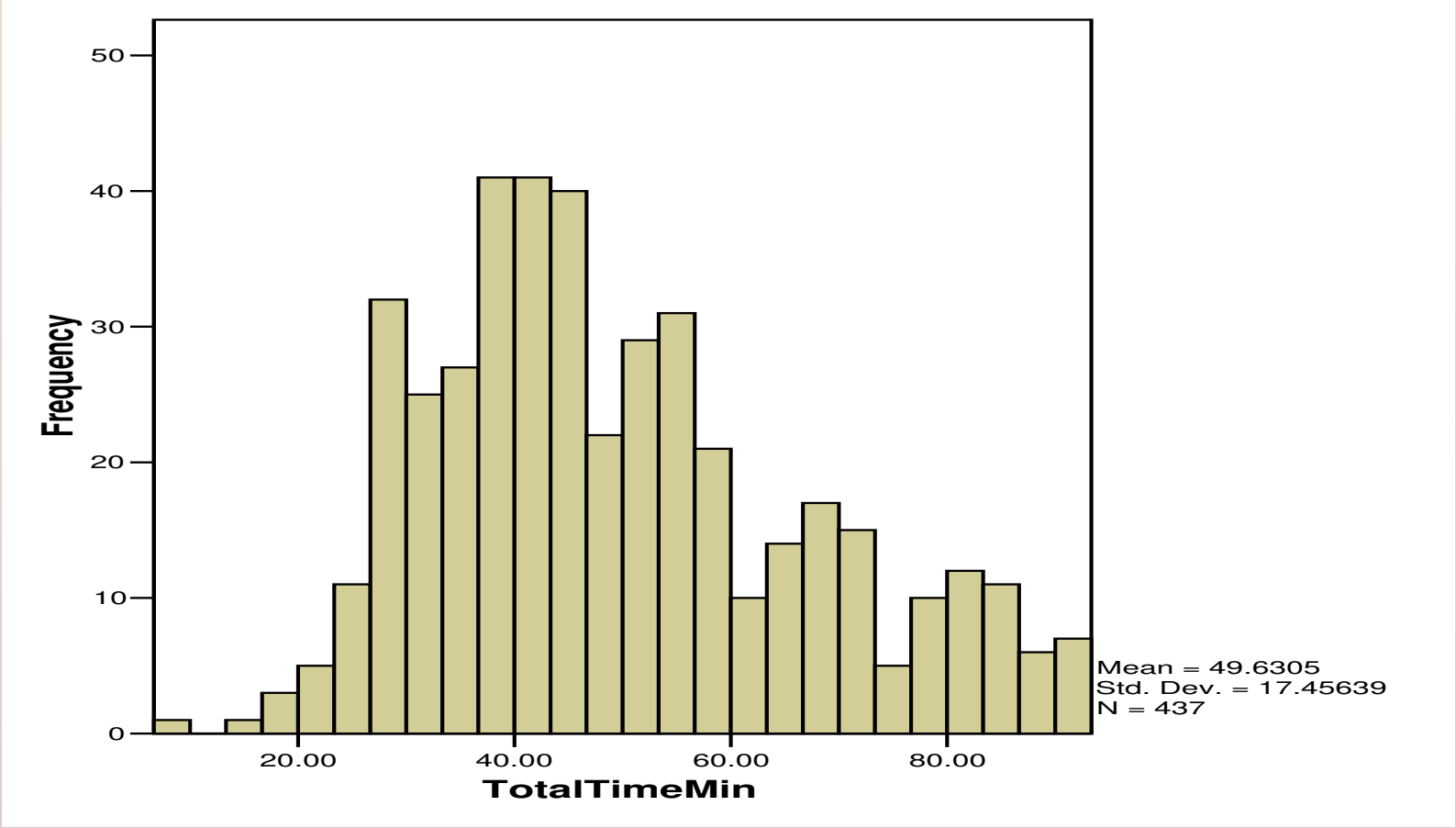
ATP Innovations in Testing 2007
Palm Springs, California

Office Email Performance Test Score Histogram



ATP Innovations in Testing 2007
Palm Springs, California

Office Email Performance Test Score Time Distribution (Minutes)



ATP Innovations in Testing 2007
Palm Springs, California

Statistics Computed

- **Classical item Statistics**
  - **P Values**
  - **Pt. Biserials**
  - **Hi-Lo Discrimination**
- **IRT Statistics**
  - **Rasch Calibration**
  - **One, Two, Three Parameter BILOG-MG Calibration**
  - **Item and Test Information**
- **Response Time**
  - **Item Time in Seconds, Test Time in Minutes, Natural Log of Test Time**

ATP Innovations in Testing 2007 Palm Springs, California

Summary of Classical Statistics

Statistic	Mean	Median	Mode	Range	Minimum	Maximum
Pt Bis	0.439	0.466	0.201	0.418	0.201	0.619
P Value	0.808	0.843	0.437	0.533	0.437	0.970
Std Dev	0.360	0.364	0.249	0.330	0.170	0.500
QValue	0.192	0.157	0.030	0.533	0.030	0.563
HiGroupP	0.913	0.950	0.963	0.473	0.527	1.000
LoGroupP	0.678	0.702	0.735	0.709	0.230	0.939
HiLoDisc	0.235	0.221	0.139	0.434	0.049	0.483

ATP Innovations in Testing 2007
Palm Springs, California

IRT Statistics Summary

Statistic	Mean	Median	Mode	Range	Low	High
Rasch b	0.000	-0.055	-1.240	4.660	-2.200	2.460
Bilob1PL a	0.676	0.676	0.676	0.000	0.676	0.676
Bilog1PL b	-1.768	-1.813	-2.809	3.932	-3.626	0.306
Bilog2PL a	0.786	0.804	1.119	1.046	0.267	1.313
Bilog2PL b	-1.700	-1.706	-3.974	4.510	-3.974	0.536
Bilog3PL a	0.863	0.872	0.305	1.176	0.305	1.481
Bilog3PL b	-1.368	-1.308	-1.071	5.123	-3.701	1.422
Bilog3PL c	0.180	0.193	0.200	0.133	0.095	0.228

ATP Innovations in Testing 2007
Palm Springs, California

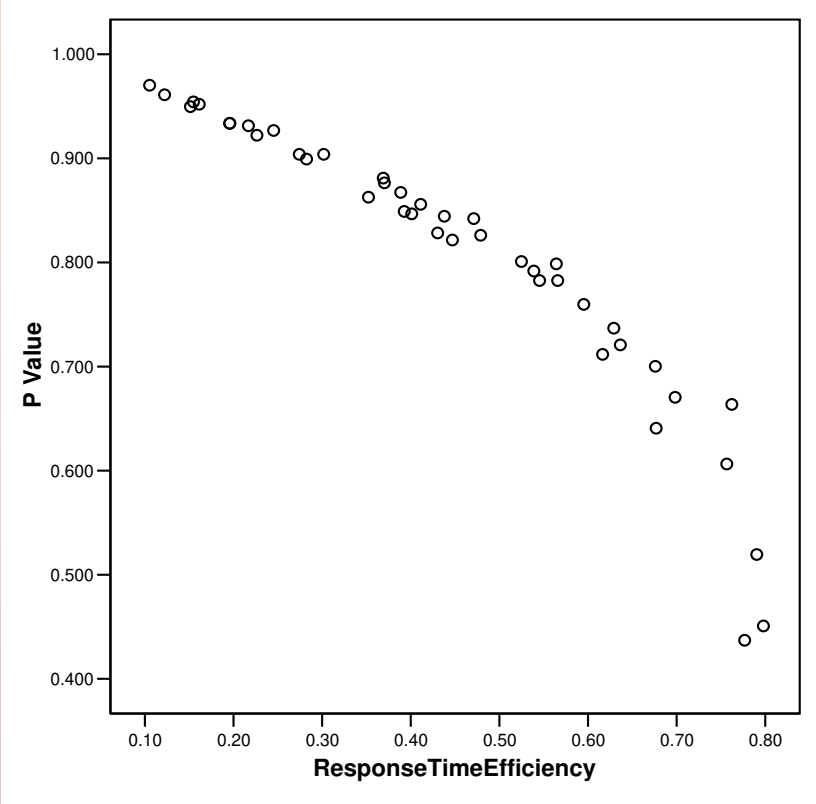
Index of Response Time Efficiency

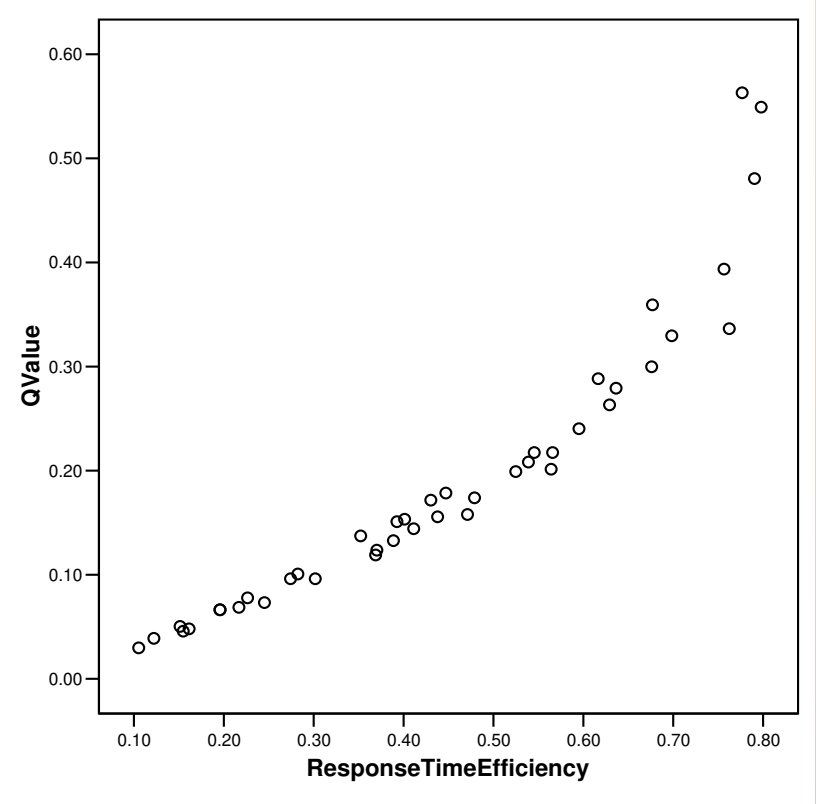
Ln((1 +(PValue*Qvalue))*π(LnTime))/LnStdTime)

Pvalue = Pvalue Classical Statistic
Qvalue = (1-Pvalue)
Pvalue*Qvalue = Item Variance, Max Value 0.25/item
π = 3.14159
LnTime = natural Log of Test Time
LnStdTime = natural Log of Time Std Dev.
Index goes from 0 to 0.80 with this dataset.

ATP Innovations in Testing 2007
Palm Springs, California

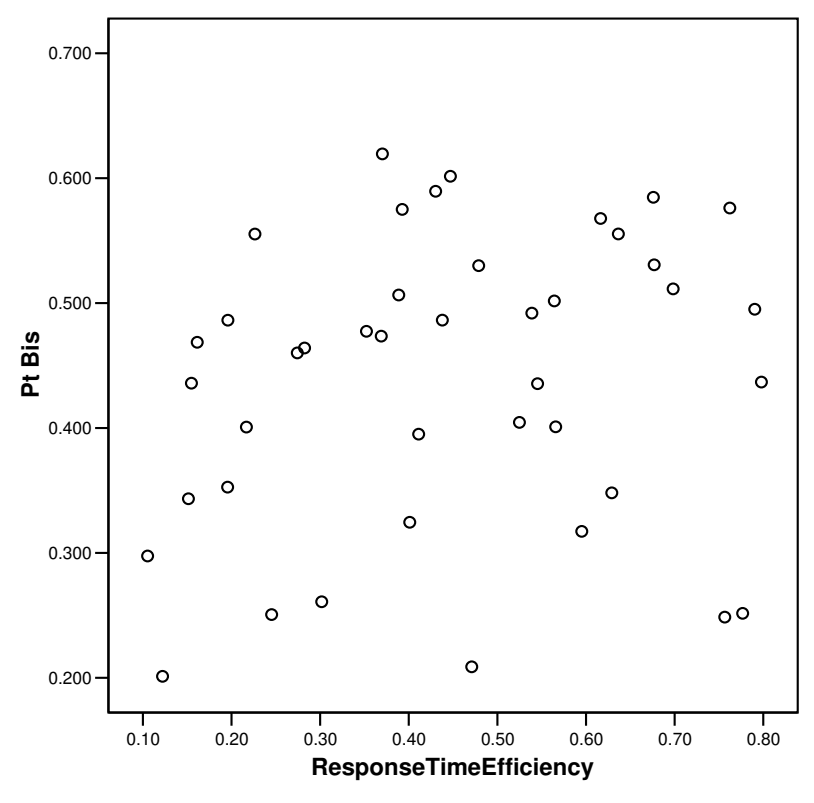
Response Time Efficiency vs. P Value Q Value

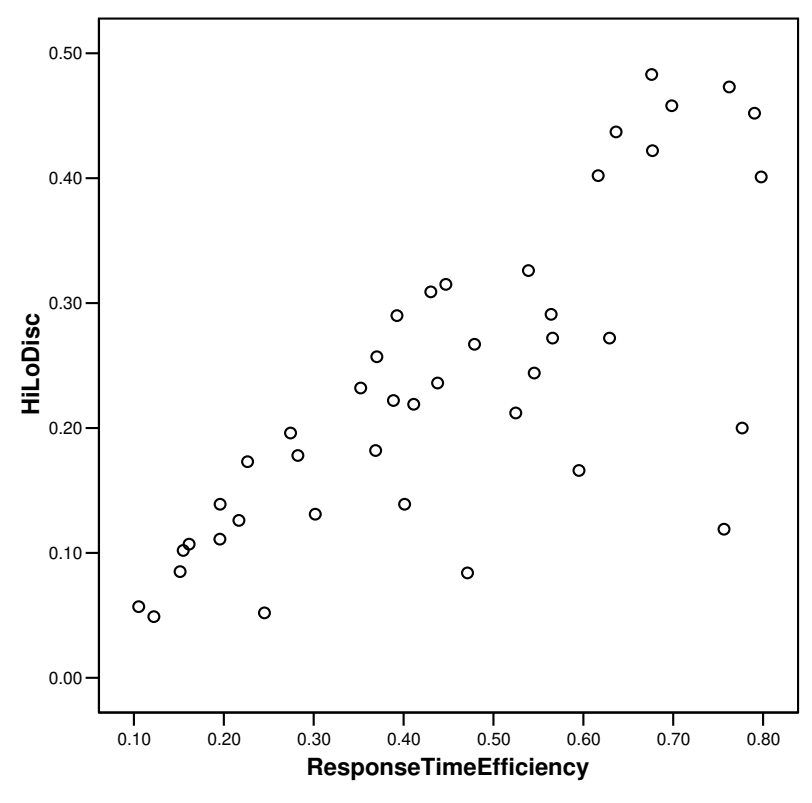




ATP Innovations in Testing 2007
Palm Springs, California

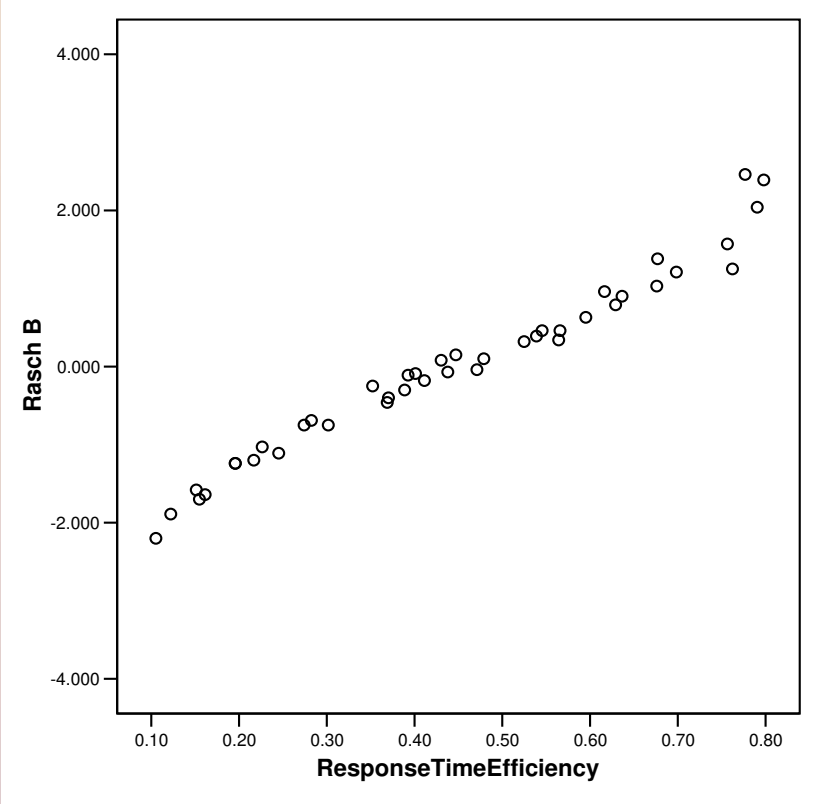
Response Time Efficiency vs. Point Biserial HiLo Disc

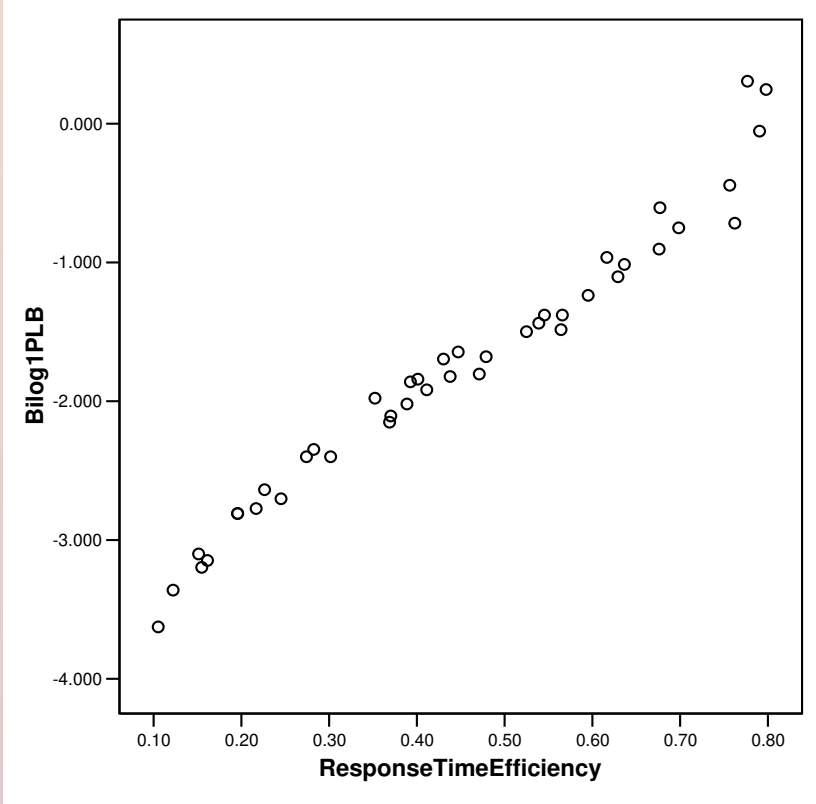




ATP Innovations in Testing 2007
Palm Springs, California

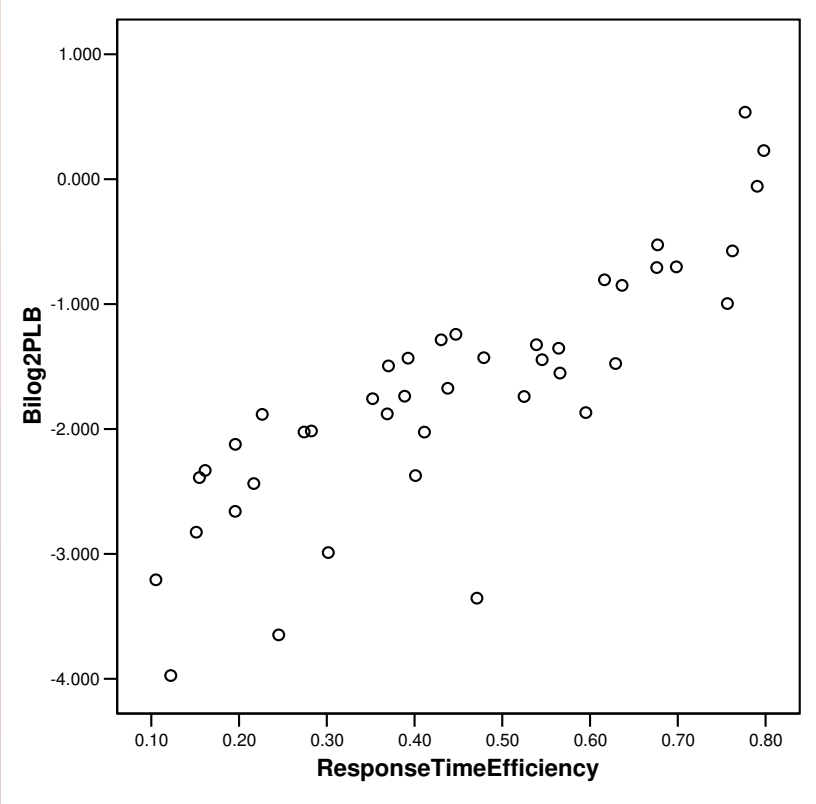
Response Time Efficiency vs. Rasch b Bilog 1Par b

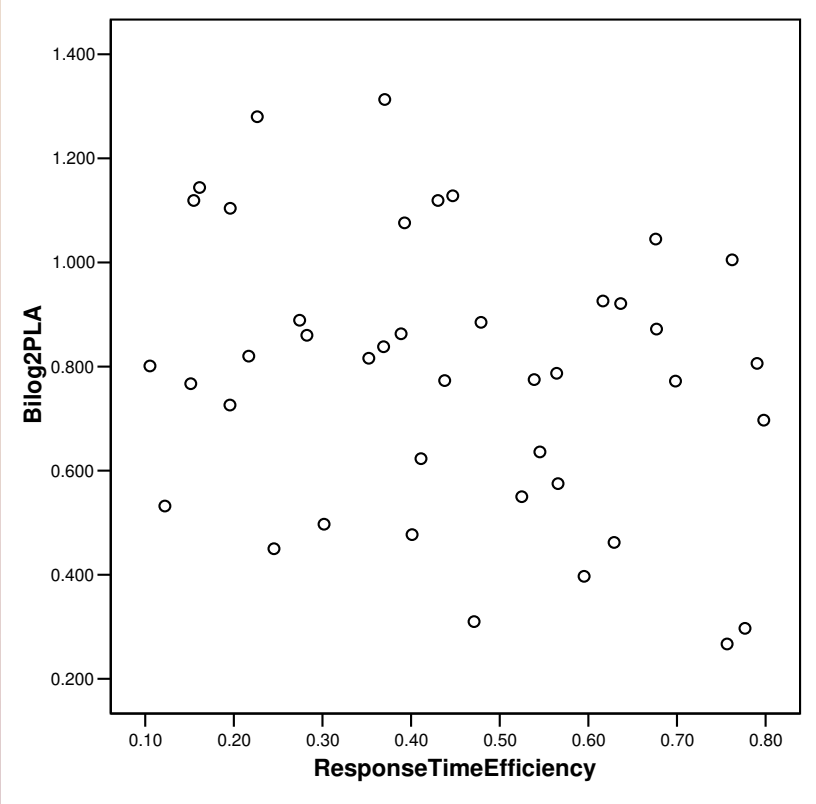




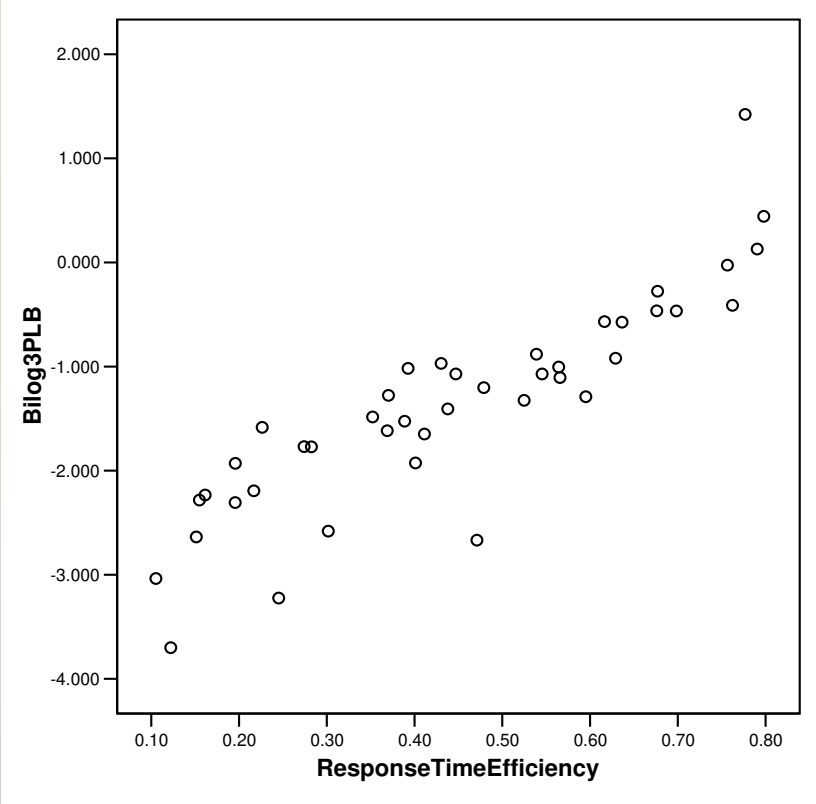
ATP Innovations in Testing 2007
Palm Springs, California

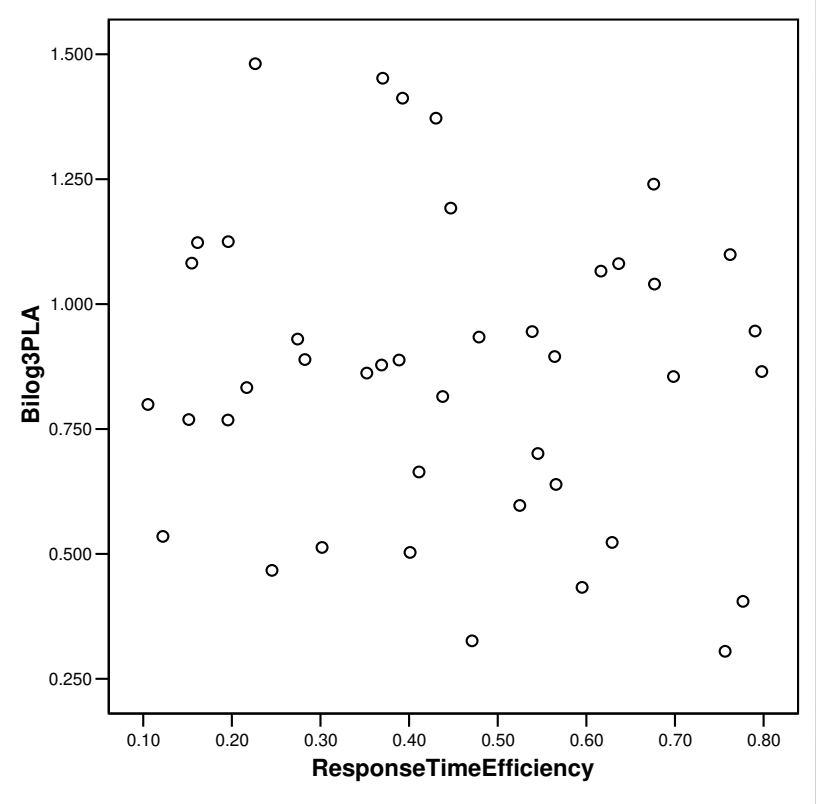
Response Time Efficiency and Bilog 2PAR b Bilog 2PAR a





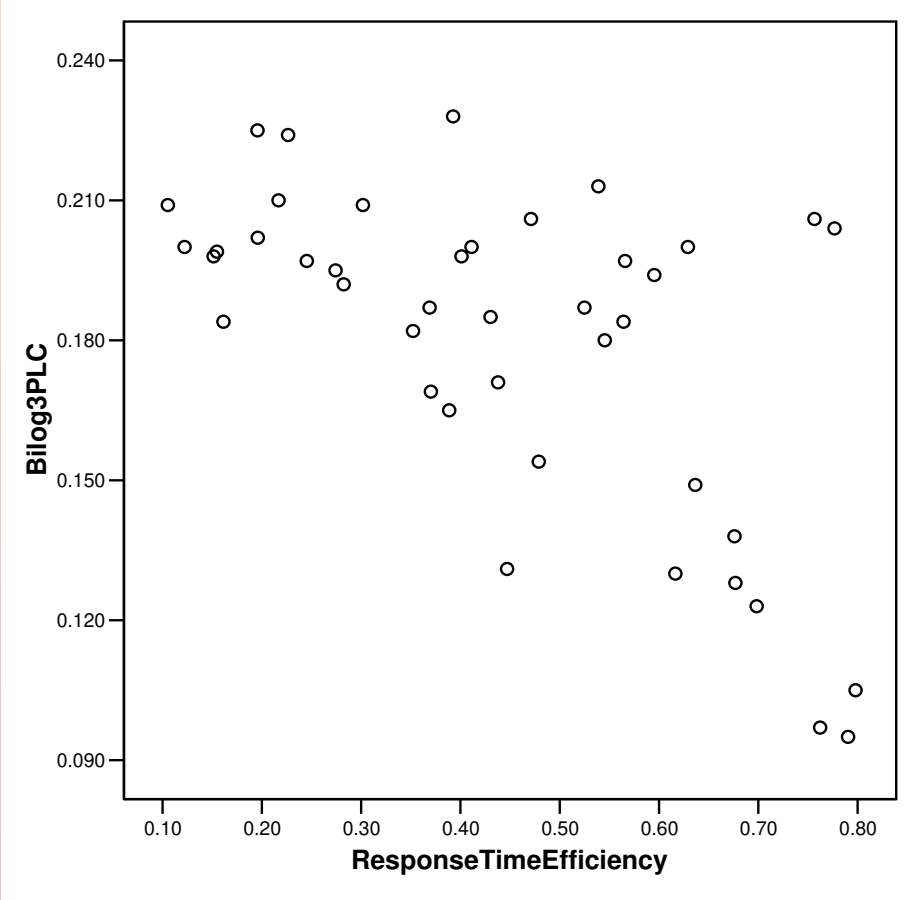
Response Time Efficiency vs. Bilog 3PAR b Bilog 3PAR a





ATP Innovations in Testing 2007
Palm Springs, California

Response Time Efficiency and Bilog 3 Parameter c value



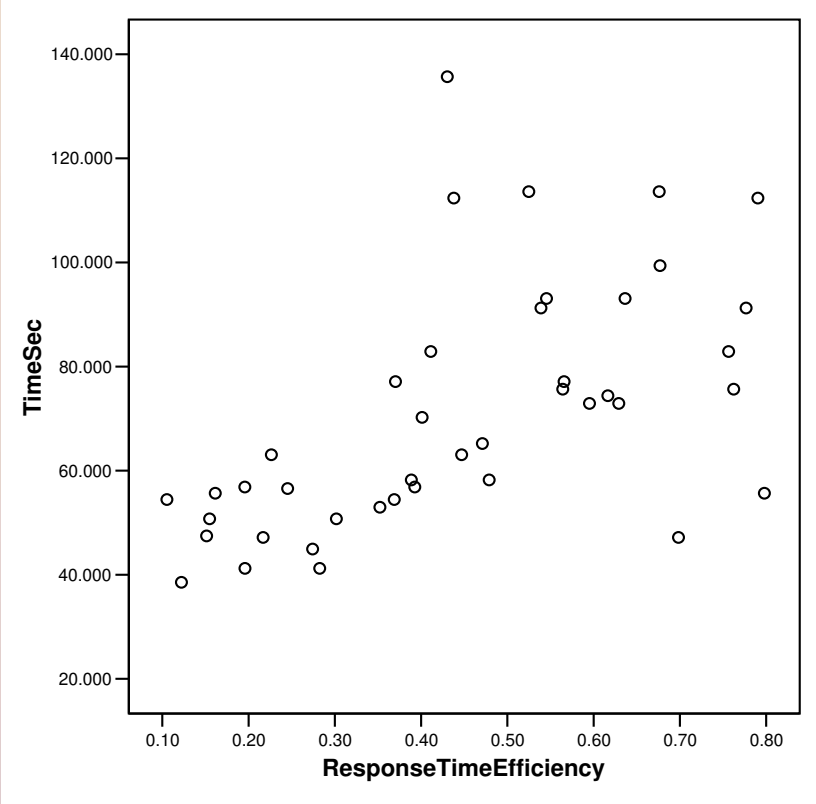
ATP Innovations in Testing 2007
Palm Springs, California

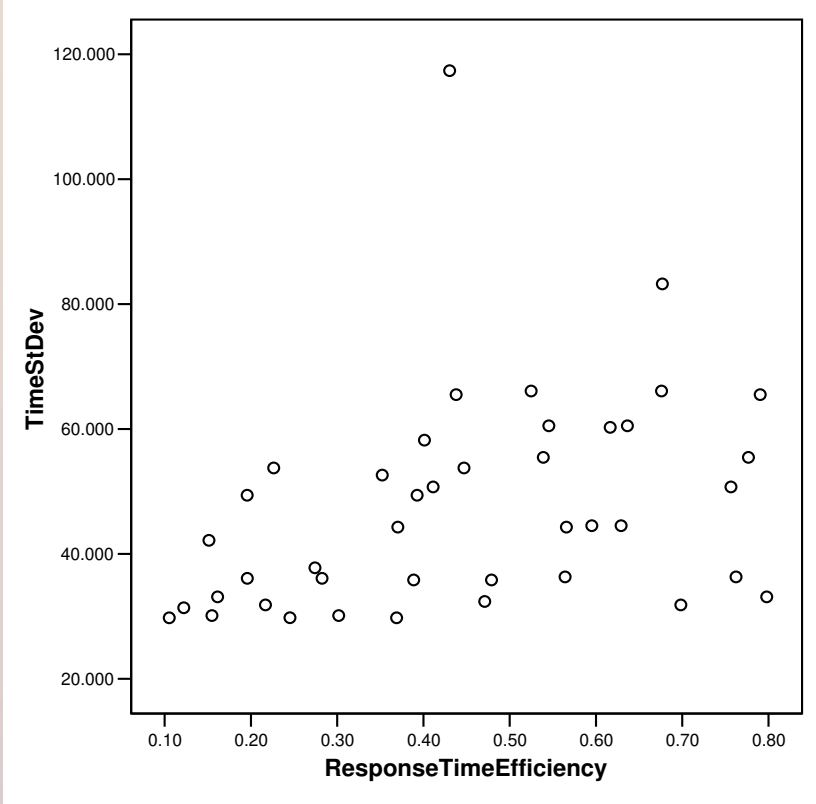
Test Response Time Summary Statistics

Statistic	N	Minimum	Maximum	Mean	Std. Deviation
TimeSec	42	38.549	135.684	70.901	23.536
TimeStDev	42	29.768	117.374	47.427	17.289
LNTime	42	3.65	4.91	4.211	.318
LNSTDTime	42	3.39	4.77	3.805	.322
ResponseTime Efficiency	42	.11	.80	.446	.206

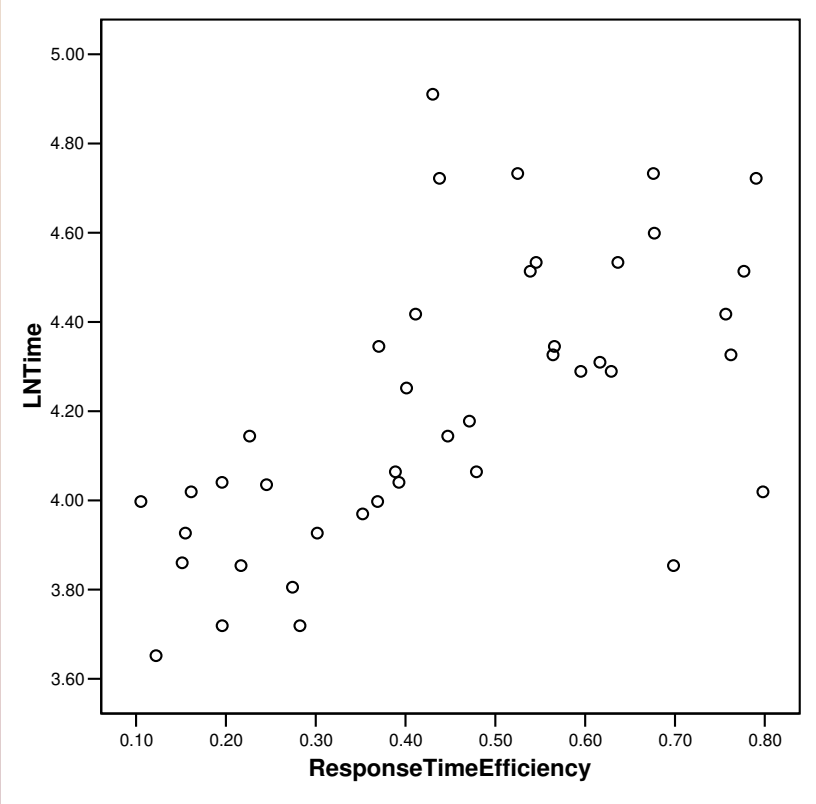
ATP Innovations in Testing 2007
Palm Springs, California

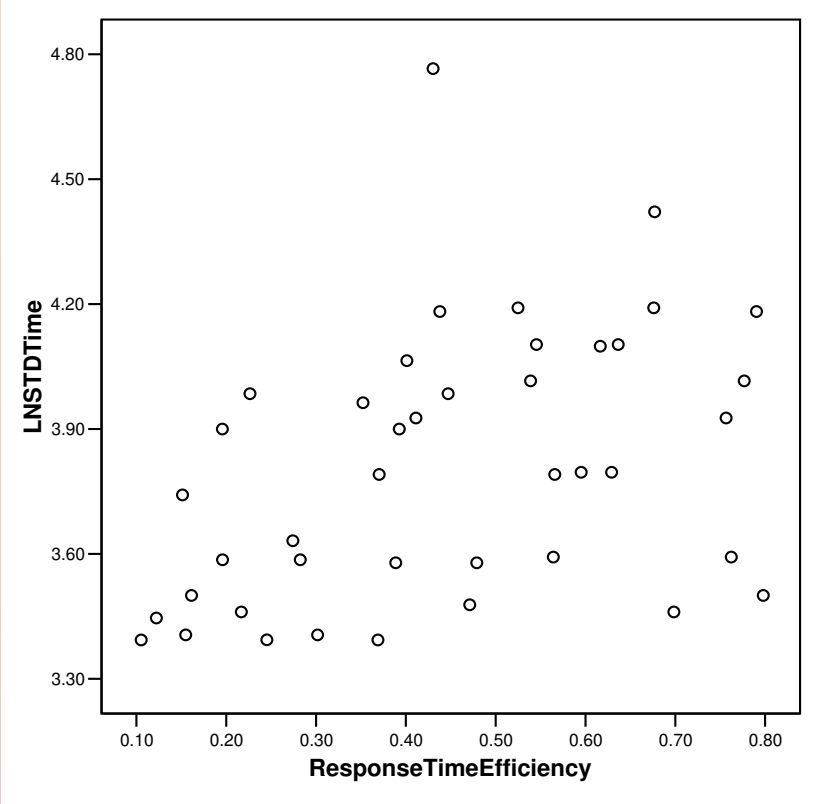
Response Time Efficiency vs. Test Time Sec Test Time StDev





Response Time Efficiency and LN Test Time LN TestTimeStDev





ATP Innovations in Testing 2007
Palm Springs, California

Response Time Efficiency Correlations with Statistics and Time

Statistic	Correlation
Pt Bis	0.20
P Value	-0.93*
Std Dev	0.99*
QValue	0.93*
HiGroupP	-0.78*
LoGroupP	-0.95*
HiLoDisc	0.76*
Rasch B	0.98*
Bilog1PLB	0.98*
Bilog2PLA	-0.27
Bilog2PLB	0.84*
Bilog3PLA	-0.12
Bilog3PLB	0.88*
Bilog3PLC	-0.61*
TimeSec	0.57*
TimeStDev	0.35
LNTime	0.62*
LNSTDTime	0.40*

ATP Innovations in Testing 2007
Palm Springs, California

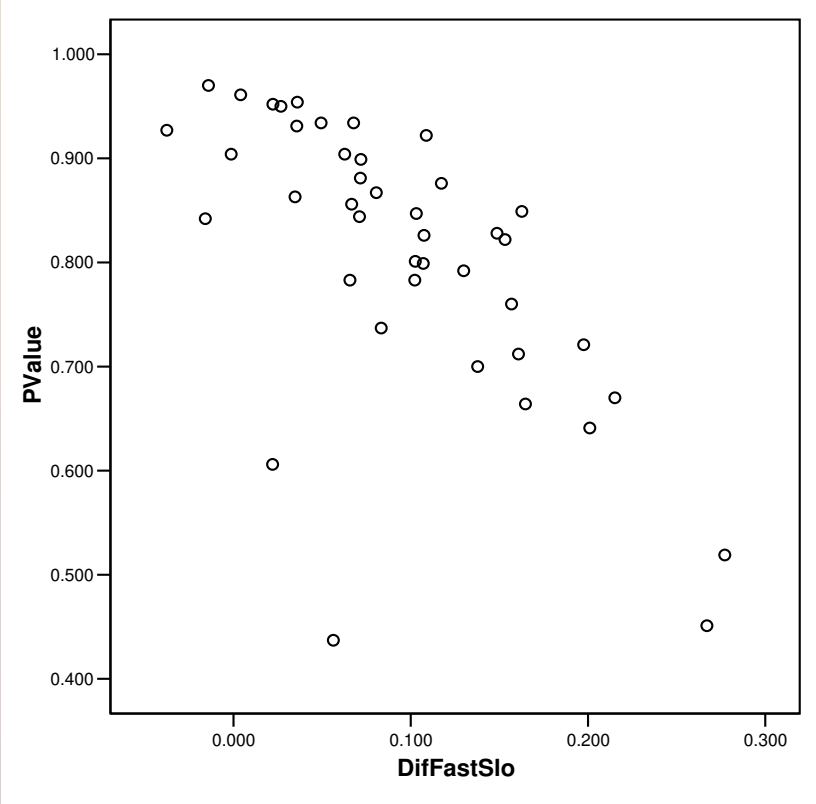
Response Time from Item Set

TotalTimeMin	Statistic
N	437.000
Mean	49.630
Std. Error of Mean	0.835
Median	45.883
Mode	48.717
Std. Deviation	17.456
Variance	304.726
Skewness	0.555
Std. Error of Skewness	0.117
Kurtosis	-0.397
Std. Error of Kurtosis	0.233
Range	81.850
Minimum	8.283
Maximum	90.133

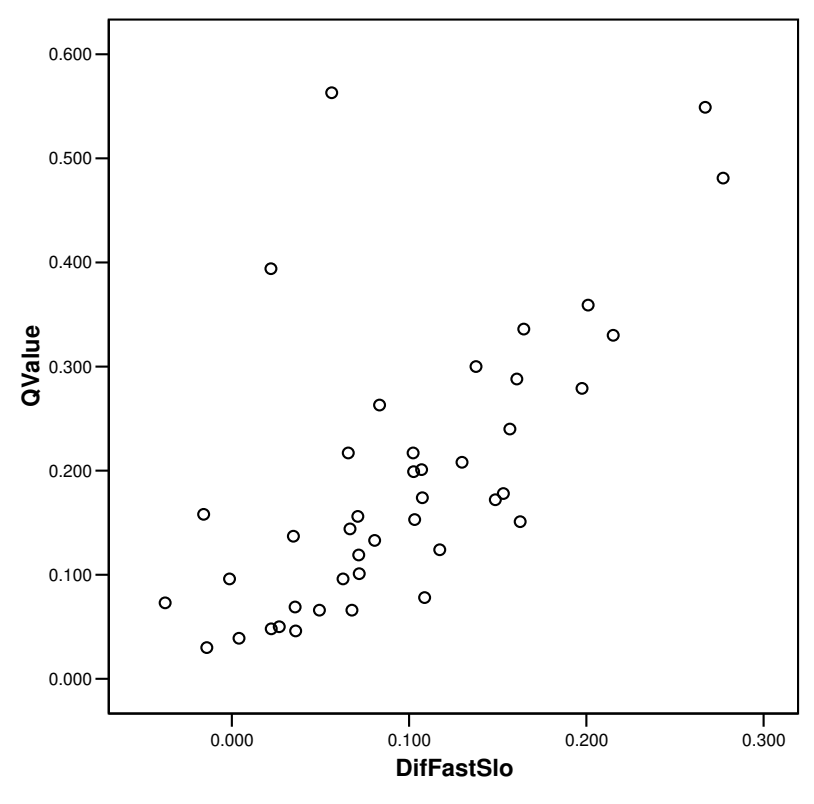
ATP Innovations in Testing 2007
Palm Springs, California

Differences Fast and Slow

PValue

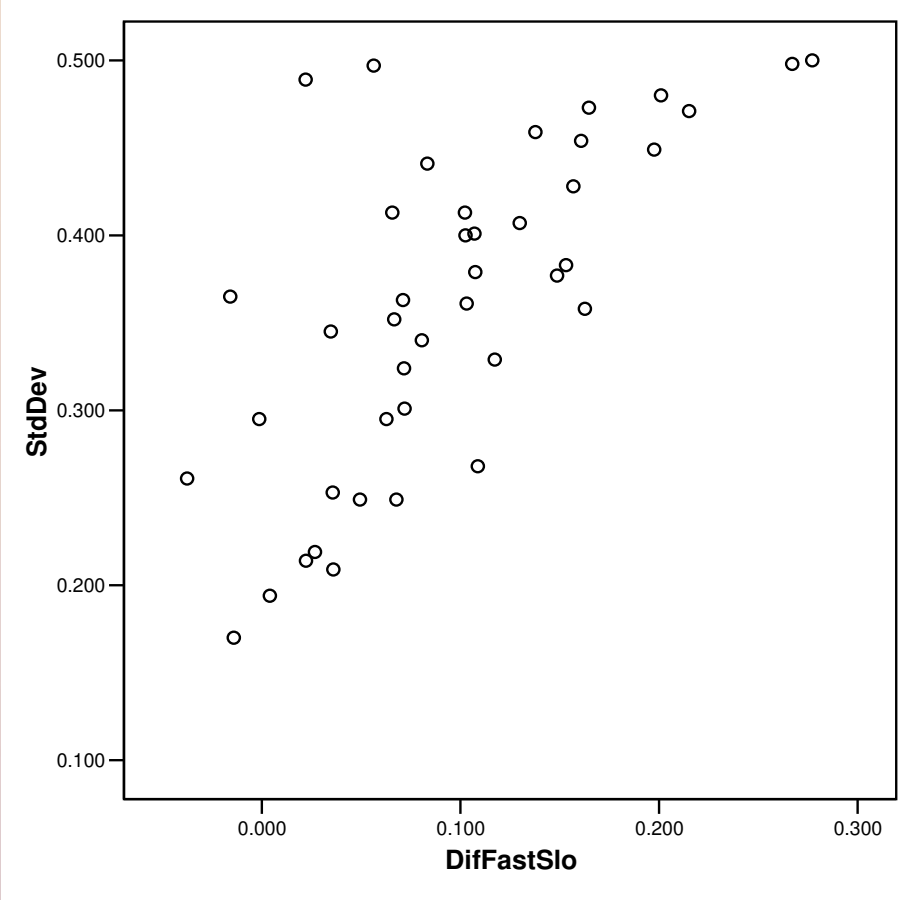


QValue



ATP Innovations in Testing 2007
Palm Springs, California

Differences Fast-Slow Item Standard Deviations



ATP Innovations in Testing 2007
Palm Springs, California

Items Showing Greatest Difference Between Slow and Fast Total Test Times

Item ID	DifPFastSlo	PValueSlow	PValueFast	Pvalue	StdDev
OL03S_C1A_102	0.277	0.382	0.659	0.519	0.500
OL03S_C1A_141	0.267	0.318	0.585	0.451	0.498
OL03S_C1A_232	0.215	0.779	0.564	0.670	0.471
OL03S_C1A_301	0.201	0.774	0.541	0.641	0.480
OL03S_C1A_062	0.198	0.820	0.623	0.721	0.449
OLS3S_C1A-202	0.056	0.409	0.465	0.437	
OLS3C_C1A-271	0.035	0.845	0.880	0.863	

Examinees were Divided at the Median Test Time in Minutes
Fast Examinees had Low Test Times,
Slow Examinees Had Long Test Times
Blue Items Fast Students Do Better
Green Items Slow Students Do Better
Grey Items No Meaningful Differences

ATP Innovations in Testing 2007 Palm Springs, California

Partial Correlations of Item Score with Item Time Controlled for Ability

Control Variables	Item ID	Correlation of Item Score to Item Time	Significance (2 Tailed)
Rasch_Ability	OL03S_C1A_062	-.199	.000
	OL03S_C1A_212	-.140	.003
	OL03S_C1A_232	-.099	.039
	OL03S_C1A_242	-.208	.000
	OL03S_C1A_271	-.165	.001
	OL03S_C1A_281	-.155	.001
	OL03S_C1A_291	-.195	.000

Investigate if Item Characteristics or Person Characteristics are responsible for the Significant Correlations Of Item Score to Item Time after controlling for Rasch Ability

ATP Innovations in Testing 2007 Palm Springs, California

Item Score Correlations to lognormal Time Controlling for Ability (Rasch Theta)

Control Variables	Item ID	Correlation of Item Score to LN(Item Time)	Significance (2 Tailed)
Rasch_Ability	OL03S_C1A_032	.115	.016
	OL03S_C1A_062	-.153	.001
	OL03S_C1A_082	.090	.060
	OL03S_C1A_102	-.276	.000
	OL03S_C1A_142	.224	.000
	OL03S_C1A_161	-.095	.047
	OL03S_C1A_171	-.272	.000
	OL03S_C1A_191	.265	.000
	OL03S_C1A_192	-.166	.001
	OL03S_C1A_212	-.161	.001
	OL03S_C1A_242	-.097	.043
	OL03S_C1A_271	-.094	.049
	OL03S_C1A_281	-.101	.035
	OL03S_C1A_291	-.149	.002
	OL03S_C1A_311	.100	.037

ATP Innovations in Testing 2007 Palm Springs, California

Index of Response Time Efficiency

Ln((3.17+(item_Std))*(LnTime))/3.17*LnStdTime)

Item_Std

LnTime = natural Log of item Time

LnStdTime = natural Log of Time Std Dev.

Index goes from 0 to 2.00 with this dataset.