Evaluation of Multi-Objective Optimizers for Cognitive Radio using Psychometric Methods

Analysis using Unidimensional and Multidimensional Rasch Models

Carl B. Dietrich
Wireless @ Virginia Tech, Bradley Department of Electrical and Computer Engineering
Blacksburg, VA, USA

Edward W. Wolfe
Pearson
Iowa City, IA, USA

Garrett M. Vanhoy
University of Arizona
Tucson, AZ, USA

***Abstract*—Item response models (IRMs) developed for use in fields such as education and psychology are applicable to cognitive radio testing due to parallels between cognitive radio and human cognition appear likely to enable efficient, and possibly adaptive testing of cognitive radios. A simulation study used unidimensional and multidimensional item response models to evaluate multi-objective cognitive engine optimizers based on two types of optimization algorithm: genetic algorithms and generalized pattern search. Data are presented in the context of cognitive radio and data are presented in a format that enables visualization of some characteristics of test items (optimization tasks) and optimizer performance identified by the IRMs. While the visualization provides intuitive confirmation of the IRM results, the IRMs identified additional significant effects that are not readily visible.**

Keywords-cognitive radio; psychometric; item response methods; Rasch measurement

I. INTRODUCTION

Efficient testing of cognitive radios is a challenging task but is important to researchers, developers, end users, and regulators. An extensive review and categorization of proposed test metrics for evaluation of cognitive radio nodes, networks, and applications is presented in [1], which also presents a testing methodology based on radio environment maps. Cognitive radios and cognitive engines are designed to include counterparts to human capabilities such as learning, decision-making, and adaptation [2, 3]. As a result, psychometric methods that have been developed for and used in testing of human cognitive characteristics have potential application to cognitive radio testing [4 (Dietrich, Wolfe, Vanhoy)]. These methods include use of item response theory (IRT) [5, 6], which comprises Rasch models and related methods and underlies adaptive testing of human cognition. IRT encompasses several methods for modeling latent traits or abilities as well as test item characteristics.

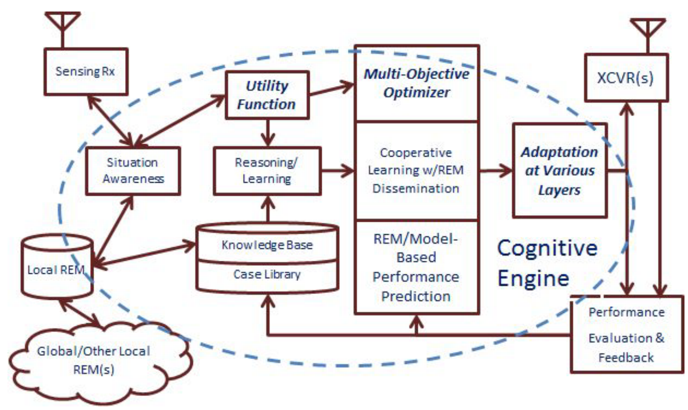


Figure 1. Cognitive Radio and Cognitive Engine (after [1]). This study focuses on the blocks labeled in italic text.

An IRT approach enables efficient test construction, and provides diagnostic capabilities for test items as well as test takers. A preliminary proof-of-concept application of IRT in [4] used a unidimensional Rasch model to measure performance of cognitive engine optimizers.

Here we describe a further application in which both unidimensional and multidimensional Rasch models are used to analyze performance of two classes of cognitive engine optimizers in optimizing multi-objective cost functions in which the relative weighting of three sub-objectives is varied. In addition to the simulation methodology, results are presented graphically to enable visualization of CE optimizer performance differences that are described by the IRT models. The remaining sections provide an overview of the unidimensional and multidimensional Rasch models used, summarize the results, and discuss implications for cognitive radio. A detailed description of the analysis is reported in [7].

CROWNCOM 2012, June 18-20, Stockholm, Sweden

DOI 10.4108/icst.crowncom.2012.248438

II. Simulation Methodology

A. Cognitive Engine Optimizers

A Multi-objective optimizer is a widely used component of a cognitive engine as shown in Figure 1 [1]. The simulation study presented here focuses on performance of the Optimizer in adapting transmitter parameters to approximate optimal performance for a variety of objective prioritizations. This serves as a useful proof-of-concept application in preparation for evaluation of the entire cognitive engine's or cognitive radio's response to its environment.

Fifty Cognitive Engine (CE) optimizers were implemented by varying parameters of two types of optimization algorithm provided in the MATLAB optimization toolbox: Genetic Algorithms (GA) and Generalized Pattern Search (GPS). These optimizers, analogous to human test takers in the IRT analysis, minimized multi-objective cost functions for cognitive radio applications that were based on those presented in [8]. The optimizers were allowed to vary three parameters of a cognitive radio: Transmitted Power P (measured in dBm), Modulation Order M (number of possible symbols, each representing $\log_2 M$ bits of information), and Frame Length L (measured in bytes). Parameters related to the environment external to the radio were fixed in these simulations, as were other radio parameters.

B. Multi-objective cost functions and test items

Thirty-six test items were designed that were intended to measure the cognitive engine optimizers' effectiveness and efficiency in optimizing multi-objective cost functions, weighted sums of individual, specialized cost functions, over a variety of weightings. Each cost function is designed so that its values range between zero and one, with the intent that when it is minimized, a specific desired performance characteristic of the radio is optimized. Test items involved three objectives: maximum good (error-free) throughput; minimum power consumption; and maximum spectral efficiency. The corresponding cost functions (see Eq. 1-3) are based on those presented in [8] and were added after multiplication by weights that ranged from 0.1 to 0.8. The sum of the three weights and hence the maximum value of the multi-objective cost function (Eq. 4) were equal to 1.0 in all cases.

$$f_{\max\_throughput} = 1 - \frac{L}{L+O+H} \cdot \left(1 - P_{be}\right)^{8(L+O)} \cdot \frac{R_s \cdot \log_2 M}{R_{s_{\max}} \cdot \log_2 M_{\max}} \cdot R_C \cdot TDD \qquad \text{(Eq. 1)}$$

$$f_{\min\_power} = 1 - \left[\begin{array}{l} \alpha \cdot \dfrac{\left(P_{mzx} + B_{\max}\right) - \left(P + B\right)}{P_{mzx} + B_{\max}} \\ + \beta \cdot \dfrac{\log_2 M_{\max} - \log_2 M}{\log_2 M_{\max}} \\ + \lambda \cdot \dfrac{R_{s_{\max}} - R_s}{R_{s_{\max}}} \end{array} \right] \qquad \text{(Eq. 2)}$$

$$f_{\max\_spectral\_efficiency} = 1 - \frac{\left(\dfrac{\log_2 M \cdot R_s}{B}\right)}{\left(\dfrac{\log_2 M_{\max} \cdot R_{s\max}}{B_{\min}}\right)} = 1 - \frac{\log_2 M \cdot R_s \cdot B_{\min}}{\log_2 M_{\max} \cdot R_{s\max} \cdot B} \qquad \text{(Eq. 3)}$$

$$f_{multi\_objective} = w_{\max\_throughput} \cdot f_{\max\_throughput} + w_{\min\_power} \cdot f_{\min\_power} + w_{\max\_spectral\_efficiency} \cdot f_{\max\_spectral\_efficiency} \qquad \text{(Eq. 4)}$$

where f denotes a cost function, w indicates a cost function weight, and subscripts identify the objective of interest. Table I lists parameters used in Eq. 1-4.

TABLE I. Parameters used in simulations. Subscripts min and max denote minimum and maximum parameter values

Parameter	Description
L	frame length in bytes (**Set directly by optimizer**)
O	physical-layer overhead (fixed at 52.5 bytes per frame)
H	MAC and IP layer overhead (fixed at 40 bytes per frame)
P_{be}	probability of bit error, or bit error rate (BER) (determined by signal power, modulation order, and channel characteristics)
R_C	coding rate (fixed at 3/4)
TDD	proportion of time during which the transmitter is active, assuming time-division duplexing (fixed at 30%)
α, β, and λ	weights that determine the relative emphasis of power and bandwidth, modulation order, and symbol rate, respectively in the cost function for minimum power consumption, where α + β + λ = 1. For the simulations presented here, α=0.5, β=0.25, and λ=0.25
P	Transmitted signal power in dBm (decibels relative to one milliwatt) **Set directly by optimizer**
B	Bandwidth of transmitted signal in megahertz
M	modulation order: number of possible symbols in the digital modulation scheme, where each symbol represents a unique sequence of $\log_2 M$ bits (**Set directly by optimizer**)
R_s	symbol rate in symbols per second (Fixed at 10^6)

This work is supported by the National Science Foundation under Grant number 0851400 and by the Institute for Critical Technology and Applied Science (ICTAS). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or ICTAS.

TABLE II. COST FUNCTION WEIGHTS AND OPTIMAL PARAMETER SELECTIONS FOR EACH TEST ITEM

Test Item	Weights used in multi-objective cost function			Optimal parameter values		
	Max. through-put	***Min. Power***	***Max. Spectral Efficiency***	**Trans-mit Power *P* (dBm)**	**Modu-lation Order *M***	**Frame Length *L* (bytes)**
1	0.1	0.8	0.1	-8	2	100
2	0.1	0.7	0.2	-8	2	100
3	0.1	0.6	0.3	-8	2	100
4	0.1	0.5	0.4	-8	2	100
5	0.1	0.4	0.5	-8	2	100
6	0.1	0.3	0.6	-8	2	100
7	0.1	0.2	0.7	-8	256	100
8	0.1	0.1	0.8	-8	256	100
9	0.2	0.7	0.1	-8	2	100
10	0.2	0.6	0.2	-8	2	100
11	0.2	0.5	0.3	-8	2	100
12	0.2	0.4	0.4	-8	2	100
13	0.2	0.3	0.5	-8	2	100
14	0.2	0.2	0.6	-8	2	100
15	0.2	0.1	0.7	-8	256	100
16	0.3	0.6	0.1	-8	2	100
17	0.3	0.5	0.2	-8	2	100
18	0.3	0.4	0.3	-8	2	100
19	0.3	0.3	0.4	-8	2	100
20	0.3	0.2	0.5	-1	4	1500
21	0.3	0.1	0.6	23	64	1500
22	0.4	0.5	0.1	-8	2	100
23	0.4	0.4	0.2	-8	2	100
24	0.4	0.3	0.3	-8	2	100
25	0.4	0.2	0.4	-1	4	1500
26	0.4	0.1	0.5	23	64	1500
27	0.5	0.4	0.1	-8	2	100
28	0.5	0.3	0.2	-4	2	1100
29	0.5	0.2	0.3	11	16	1500
30	0.5	0.1	0.4	23	64	1500
31	0.6	0.3	0.1	-4	2	1100
32	0.6	0.2	0.2	23	64	1500
33	0.6	0.1	0.3	23	64	1500
34	0.7	0.2	0.1	23	64	1500
35	0.7	0.1	0.2	23	64	1500
36	0.8	0.1	0.1	23	64	1500

Table II lists objective weights and the radio parameter values that result in a global minimum of the multi-objective cost function for each test item. The sum of the three weights is always one, and the resulting patterns in the weights are evident in Figure 2.

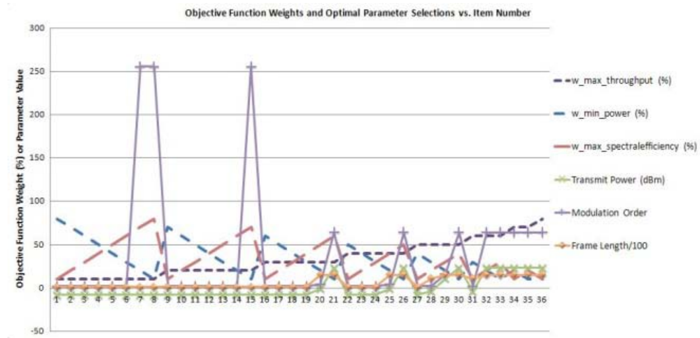


Figure 2. Objective function weights and optimal parameter selections vs. item number

TABLE III. CORRELATIONS OF OPTIMAL PARAMETER VALUES WITH OBJECTIVE FUNCTION WEIGHTS

Weights / **Parameters**	***Max. Throughput***	***Min. Power***	***Max. Spectral Efficiency***
Transmit Power *P* (dBm)	0.718	-0.622	-0.0964
Modulation Order *M* (symbols)	-0.0601	-0.515	0.575
Frame Length *L* (bytes)	0.743	-0.643	-0.0999

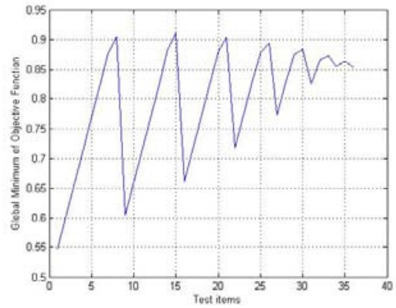
Correlations of the radio parameter values that yield an optimum solution and the weights of each objective are given in Table 3. Note that these correlations may reflect indirect effects in that if one objective is emphasized because its weight is relatively large, the others are deemphasized because the sum of the weights is fixed at one. An interpretation follows:

1) *Transmit power:*

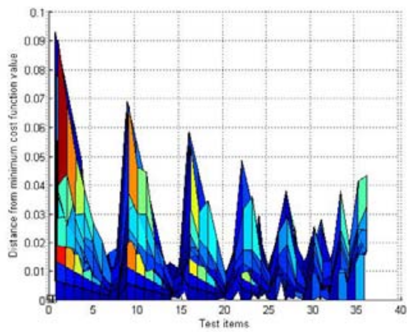
- Positive correlation with $w_{max\_throughput}$ is expected because increasing power decreases bit-error rate (P_{be}), allowing higher good (error-free) throughput
- Negative correlation with w_{min_power} is expected: Decreasing transmit power minimizes power consumption
- Near-zero correlation with $w_{max\_spectral\_efficiency}$ is expected: Transmitted power is not included in $f_{max\_spectral\_efficiency}$.

2) *Modulation order:*

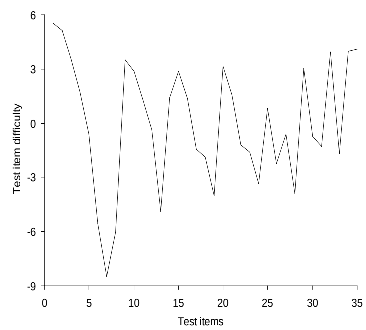
- Near-zero correlation with $w_{max\_throughput}$: increasing modulation order for fixed transmit power increases data rate but also increases BER
- Negative correlation with w_{min_power}: when w_{min_power} is high, transmitted power tends to be low, which means P_{be} at high modulation orders will be high.
- Positive correlation with $w_{max\_spectral\_efficiency}$: increasing modulation order *M* increases spectral efficiency. Bit errors are not included in $f_{max\_spectral\_efficiency}$.



(a)



(b)



(c)

Figure 3. (a) Minimum achievable value of multi-objective function for each test item. (b) distance of CE optimizer solution from optimum (minimum) value of cost function. (c) item difficulties indicated by IRM

3) Frame Length:

- Positive correlation with $w_{max\_throughput}$: Long frame length minimizes overhead as a proportion of total data transmitted, which tends to increase throughput.
- Negative correlation with w_{min_power}: As w_{min_power} increases, transmitted power will be decreased. For low transmitted power and hence low signal-to-noise ratio, P_{be} is high and long frame lengths result in low throughput. Also as w_{min_power} increases, $w_{max\_throughput}$, which emphasizes the benefit of long frame length for throughput, decreases.
- Near-zero correlation with $w_{max\_spectral\_efficiency}$: Frame length is not considered in this cost function.

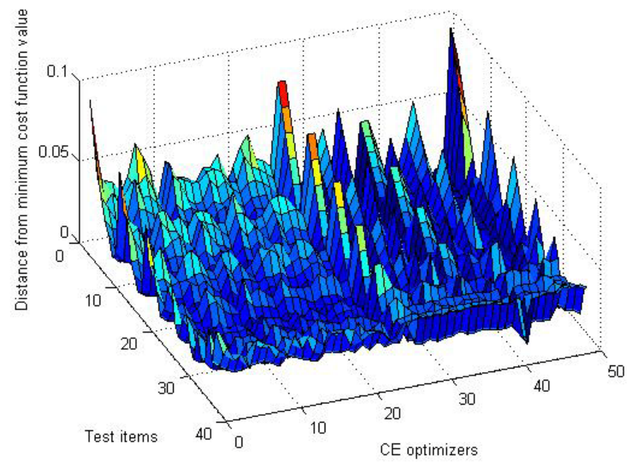


Figure 4. Absolute value of difference between minimized multi-objective cost function value and global minimum, plotted vs. Item number and Cognitive Engine Optimizer number (note optimizers 1-25 use Genetic Algorithms, 26-50 use Generalized Pattern Search.

As a result of constraints on radio parameter values and environmental characteristics, the achievable minimum of the multi-objective function varied by test item, as shown in Figure 3. Items with higher achievable minimum objective function values do not necessarily present the most difficult problems for the optimizers and in fact appear to result in CE solutions for the multi-objective cost function that are close to the desired, minimum value. However, in these cases the optimum cognitive radio performance would not approach the radio's ideal performance.

C. Effectiveness Measurement

CE optimizer effectiveness is measured using the difference between the multi-objective cost function value that results from the optimizer's parameter selections and the global minimum of the multi-objective cost function within the allowable parameter selections. The two effectiveness metrics used are the absolute difference and the difference normalized by the global minimum of the multi-objective cost function. Figures 4 and 5 present these two metrics plotted vs. Item number and Cognitive Engine number. Items were described in Table I and Figure 2. Note that CEs 1-25 use Genetic Algorithms (GA), and 26-50 use Generalized Pattern Search (GPS). Differences between the two optimizer types are apparent in the data, although some items seem to have been relatively easy for all the optimizers.

III. Item Response Models and Analysis Results

Item response models, item response theory, and their potential benefits for cognitive radio testing are discussed in [4] and are overviewed briefly here. This is followed by results obtained by using a variety of IRMs to analyze data from the simulations described in Section II. Details of the analysis itself are presented in [7].

A. Overview of Item Response Models

Item response models (IRMs) represent observable behaviors as expressions of one or more latent traits. In the case of a cognitive engine optimizer, possible latent traits include adaptation effectiveness or speed. These traits are measured by observing behaviors in response to test items and coding them into dichotomous (e.g. correct/incorrect) or polytomous data (data having more than two possible values). Polytomous data may be categorical or ordinal. Item response theory (IRT) encompasses IRMs, algorithms for estimating parameters for the models, approaches for measuring model-data fit, techniques for constructing tests that provide desired information with a minimum number of items, methods for detecting differential item functioning (a possible indicator of biased test items), ways of equating results from tests containing different sets of items. IRT applications include adaptive testing, in which test items are selected and administered dynamically based on test-takers' responses to previous items [5].

B. Examples of IRMs

To illustrate the concept of IRMs, three types of IRMs are described briefly that are suitable for use with: (1) dichotomous data; (2) polytomous data; and (3) sets of test items that measure multiple performance characteristics not sufficiently correlated to be represented by a single latent trait. Also, as mentioned earlier, IRMs can be used to identify differential item functioning (DIF).

1) Unidimensional, Dichotomous Rasch Model

This model was one of the first IRMs and is described in [9]. The model includes a single dimension or latent trait and is applicable when only two possible responses to each item are allowed or considered (e.g. true/false questions or a pass/fail threshold applied to questions that have a continuum of possible responses). In the context of testing in which items have a correct and an incorrect response, this model is used to measure levels of a single latent trait in test takers and to jointly measure the level of this trait that is required for a test taker to have a 50% probability of a correct response to a particular item. The model can be written as

$$\ln(\pi_{X=1}/\pi_{X=0})=\theta_n-\delta_i \qquad \text{(Eq. 5)}$$

or alternatively as

$$\pi_{X=1}=\exp(\theta_n-\delta_i)/[1+\exp(\theta_n-\delta_i)] \qquad \text{(Eq. 6)}$$

where $\pi_{X=1}$ is the probability of a correct response, $\pi_{X=0}$ is the probability of an incorrect response, θ_n is the level of the latent trait in the n^{th} test taker, and δ_i is the difficulty of the i^{th} item. The probability of a correct response is 0.50 if $\theta_n=\delta_i$, and becomes greater as θ_n is increased or δ_i is decreased.

2) Partial Credit (Polytomous) Model

IRMs can also accommodate polytomous ordinal outcomes by including an additional parameter that indicates the relative difficulty of obtaining a score in one rating category versus the next lower rating category. Such a model is called the partial credit model, which contains a separate item-by-threshold difficulty for each adjacent pair of score categories (δ_{ij}) [6]. The model can be written as

$$\ln(\pi_{X=k}/\pi_{X=k-1})=\theta_n-\delta_i-\tau_{ij} \qquad \text{(Eq. 7)}$$

where $\pi_{X=k}$ is the probability of a response in category k, $\pi_{X=k-1}$ is the probability of a response in the next lower category, and τ_{ij} is the relative difficulty of categories k versus k-1 on the i^{th} item.

3) Differential Item Functioning (DIF)

Dichotomous and partial credit IRMs rely on the assumption that the item difficulties are invariant across groups of test takers. That is, these models require that items exhibit the same levels of difficulty regardless of test taker. The plausibility of this assumption can be evaluated by including an additional parameter in the IRM that depicts the interaction between test taker groups and item difficulty. The model can be written as

$$\ln(\pi_{X=k}/\pi_{X=k-1})=\theta_n-\delta_i-\gamma_g-\tau_{ij}-\iota_{ig} \qquad \text{(Eq. 8)}$$

where γ_g is the ability of test takers in group g, relative to the population of test takers, and ι_{ig} is the deviation of the difficulty of item i for that group from the population difficulty of that item, denoted δ_i. By convention, large item-by-group interaction terms suggest the existence of differential item functioning and indicate that groups of test takers define the underlying latent trait in different ways

4) Multidimensional IRMs

Each of these models can be extended to take into account multiple dimensions or latent traits via the Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM). [10]. This model maps each item onto one of several latent dimensions, which are depicted by a vector of latent traits rather than the singular latent traits contained in Equations 5, 6, 7, and 8.

5) Parameter Estimation

Parameters for IRMs are estimated, typically via maximum likelihood procedures, based on observed scores for test takers on a set of test items. Once test taker and test item parameters are estimated, model-to-data fit can be evaluated both globally and separately for each test taker and item. Global fit analyses focus on identifying a best fitting model while test taker and item fit analyses focus on identifying anomalies for further study. In addition, the reliability of subscale measures and correlations between latent dimensions may be evaluated to determine the usefulness of the multiple measures obtained in multidimensional IRMs. Relative item difficulties, conditioned on known item features, may provide information about causal relationships between those item features and test taker performance. Finally, differences between the average measures of groups of test takers may provide information about differential performance of subpopulations of test takers.

C. *Data Analysis and Results*

Analysis of data from the simulations in Section II is described briefly and results are summarized. A more detailed discussion of the data analysis itself is provided in [7].

Thirty five of the items were used in the analysis. Item 14 in Table II was omitted from the analysis after it was determined that this item did not provide useful information about relative capabilities of the optimizers, due primarily to the relatively low difficulty that the simulated CEs had identifying an optimal parameterization for this item. This resulted in a full data matrix containing 50x35=1750 data points.

Effectiveness measures were transformed to three-point ordinal scores, with low scores indicating high effectiveness. Data-model fit was assessed using partial credit forms of four IRMs: unidimensional (1D), unidimensional differential item functioning (1D-DIF), two dimensional (2D), and two dimensional differential item functioning (2D-DIF). The 2D-DIF model exhibited the best fit to the observed data, followed by the 1D-DIF and 2D models, with the 1-D model providing the worst fit. Tasks in which the objective function included relatively high weight for the maximum throughput objective combined with low weight for power appeared to test a second dimension of performance from the other tasks. The test items associated with each dimension in the two 2D models are as follows:

- Dimension 1: Items 1-13, 15-19, 22-24, 27
- Dimension 2: Items 20-21, 25-26, 28-36
- Item 14 was excluded from the analysis due to nearly uniform performance by all optimizers on this task.

D. *Results*

Sample results from the analysis are shown here. The 2D-DIF model was shown to be appropriate for the data, yielding high reliability of separation of 0.97 and 0.89, respectively for each of the two latent trait dimensions identified; the correlation of Dimension 1 vs. Dimension 2 was 0.65 [7].

Table IV shows the mean and standard deviation of two latent traits (Dimensions 1 and 2) for the CE optimizers by algorithm category. Performance of the two algorithms is comparable in the second dimension, but the algorithms exhibit different functioning in the first dimension, measured by items that put a relatively lower weight on the objective of maximum throughput. The difference between categories of cognitive engine optimizers in this second dimension may be due to tradeoffs inherent in selecting modulation order and frame length to minimize a cost function that emphasizes good (error-free) throughput. The analysis reveals higher means (indicating poorer performance) and higher standard deviations for GPS algorithms (optimizers 26-50) for the two sets of test items, and while these differences are not necessarily obvious in Fig. 4, the raw data for the two types of cognitive engines do appear to exhibit different patterns in the figure, with some relatively large values evident in the data for the GPS algorithms.

TABLE IV. SUMMARY OF CE MEASURES ON TWO DIMENSIONAL MEASURES [7]. GA = GENETIC ALGORITHM. GPS = GENERALIZED PATTERN SEARCH. N = 25 FOR EACH GROUP. $T_{DIM\,1} = 3.73$, $df_{Satterthwaite} = 31.64$,

$p = .001$. $T_{DIM\,2} = 1.49$, $df = 48$, $p = .14$.

Dimension	Statistic	GA	GPS
1	Mean	−3.49	1.85
	SD	2.69	6.65
2	Mean	-0.67	0.28
	SD	1.90	2.57

IV. CONCLUSION

Following preliminary work in [4], the simulations and analysis described here further demonstrate applicability of IRMs to evaluation of cognitive radios and are a step closer to use of IRMs to enable efficient cognitive radio testing. This study evaluated ability of GA and GPS-based CE optimizers to minimize multi-objective cost functions by setting cognitive radio parameters, and shows that the two types of cognitive engines tend to perform differently for certain combinations of objective weights. Results of a detailed multidimensional IRM analysis presented in [7] are seen to be consistent with observable features in plots of raw data and also provide useful results not available by inspection of the data.

ACKNOWLEDGMENT

Thanks to Professors Tamal Bose, Tonya Smith-Jackson, and Kay Thamvichai for their help in making this work possible.

REFERENCES

[1] Y. Zhao, S. Mao, J.O. Neel, and J.H. Reed, Performance Evaluation of Cognitive Radios: Metrics, Utility Functions, and Methodology, *Proceedings of the IEEE*, Vol. 97, No. 4, pp. 642-659, April 2009.

[2] J. Mitola III, *Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio* [Dissertation], Royal Institute of Technology (KTH), Stockholm, Sweden, May 8, 2000.

[3] S. Haykin, Cognitive radio: brain-empowered wireless communications, *IEEE JSAC*, vol. 23, no. 2, pp. 201-220, Feb. 2005.

[4] C.B. Dietrich, E.W. Wolfe, G. Vanhoy, “Cognitive Radio Testing using Psychometric Approaches,” *SDR’11/WinnComm*, Washington, DC, Nov. 29-Dec. 2, 2011.

[5] R.K. Hambleton, H. Swaminathan, H.J. Rogers, *Fundamentals of Item Response Theory*,

[6] B.D. Wright, G.N. Masters, *Rating Scale Analysis: Rasch Measurement*, MESA, Chicago, IL, 1982.

[7] E.W. Wolfe, C.B. Dietrich, G. Vanhoy, “Application of the Rasch model to measureing the performance of cognitive radios,” *Fifth International Conference on Probabilistic Models for Measurement*, Perth, Australia, January 23-25, 2012.

[8] T.R. Newman, *Multiple Objective Fitness Functions for Cognitive Radio Adaptation*, Ph.D. dissertation, University of Kansas, 2008.

[9] G. Rasch, Probabilistic Models for some Intelligence and Attainment Tests, University of Chicago, Chicago, IL, 1980.

[10] R.J. Adams, M. Wilson, W.C. Wang, “The multidimensional random coefficients multinomial logit model,” *Applied Psychological Measurement*, vol. 21, pp. 1-23.