Leveraging Fidelity Data to Making Sense of Impact Results:
Informing Practice through Research

Andrew P. Jaciw

Thanh Nguyen

Empirical Education Inc.

Background. When a randomized experiment fails to demonstrate program efficacy, the obvious and important question is 'why not?' Critical to answering it is assessing levels of fidelity of implementation (FOI) and identifying how the program was impeded and may be improved. A related question is whether impact is observed under conditions of adequate FOI. Fidelity can be defined variously (Hulleman and Cordray, 2009) in terms of the processes and stages of the logic model, including levels of: delivery of program inputs, uptake of inputs (Participation), Receptiveness to the program supports and professional development, and Adherence to program principles and practices.

A challenge to assessing impact under strong implementation is the potential for bias from endogeneity. While randomization yields unbiased estimates of marginal impact (barring high attrition), limiting the sample to strong implementers in the treatment condition leads to the problem of identifying an equivalent group in control. Analysis of Symmetrically Predicted Endogenous Subgroups (ASPES) (Peck, 2003; Peck, 2016; Unlu et al. 2010) is an approach that circumvents this problem by identifying strong implementers in both conditions in terms of baseline characteristics, and assessing impact within subgroups, thus identified. Critically, it provides insight into aspects of practice, including modifiable factors, consequential for impact.

Purpose: This work has two main purposes: (1) to present a variant of ASPES elaborated on in Unlu et al (2010) that assesses whether impact varies across a continuous measure of FOI that is predicted from rich baseline data; (2) to show how result of this type of ASPES analysis varies depending on the definition and scaling of FOI; specifically, we compare results of more-superficial attendance-based measures of FOI (Participation, Receptiveness) to deeper Adherence- and practice-based measures, and show that they perform differently, with the latter conceptualization shedding insight on reasons for a lack of gross impact, including variations in malleable aspects of practice.

Research Design / Setting / Subjects / Intervention: The study evaluated the efficacy of the internet-based Reading Apprenticeship Improving Science Education (iRAISE) intervention. iRAISE is a high-school literacy program implemented in science classes, that uses meta-cognitive strategies for advancing reading. Eighty-two teachers and 1,468 students were randomly assigned to iRAISE or Business as Usual. Students' reading literacy was assessed after one year. The study took place in rural Pennsylvania and Michigan.

Data Collection and Analysis:
Defining FOI. FOI was defined in three ways, as: (1) "Participation" (e.g., participation in professional training), (2) "Receptiveness" (e.g., perceived helpfulness of components of program), and (3) "Adherence" (e.g., demonstrating principles of iRAISE in instruction). (Data

in (1) and (2) were from teacher surveys, and (3) from facilitators of Professional Learning Communities). Importantly for this work, 1 and 2 are more-superficial indicators of take-up of iRAISE, while 3 gets at core practices. Indicators in each of (1) and (2) were combined multiplicatively, or additively (i.e., in a compensatory way); (3) was reported holistically, and in terms of multiple dimensions that were averaged. This resulted in six metrics total (see Appended Table 1.)

ASPES method. The brevity of this abstract precludes a detailed description of the method (see Unlu et al., 2010). In short, it involves three steps: (1) Regress FOI against baseline characteristics (BC) in the treatment group, (2) use the model to produce predicted values of fidelity (FOI*) for teachers in both conditions, (3) assess the interaction between FOI* and the dummy variable for treatment assignment. (See Appended Figure 1)

Modeling FOI to get predicted values (FOI)*: We used three approaches to choose baseline covariates for predicting FOI. A: forcing in literacy-related items and forward selecting from among demographics; B: forward selecting from among all available covariates; Theoretical: Choosing only covariates that theory and commonsense dictate should influence FOI for this intervention.

The alternatives above resulted in 18 analyses (6 metrics of FOI × 3 prediction models).

Findings / Results:

Adequacy of prediction models: Figure 2 shows that the predicted FOI metrics reflect observed FOI: Close to 80% of the variance in FOI scores is accounted for by baseline characteristics (except in the Theoretical model for the "Participation" and "Receptiveness" FOI metrics). Also, the correlation between observed levels of fidelity (FOI) and predicted levels of fidelity (FOI*) was close to .90 (lower for the Theoretical Model for "Participation" and "Receptiveness"). This gives assurance of the validity of model-predicted values of fidelity to the iRAISE program.

Average impact: There was no impact of iRAISE on reading literacy on average (standardized effect size was .002, $p = .96$)

Relationship between FOI and impact on achievement: Results are summarized in Figure 3. (Full results in Table 2). We observe that the Adherence measures of FOI have higher *t* values compared to the other measures of FOI, and two – Average (Model B) and Holistic (Model A) – each reach marginal statistical significance ($p<.10$). The added value of impact, on student science literacy achievement, for a 1-unit increase in FOI on these two scales is .04 and .03 standard deviation units, respectively. The remaining FOI measures do not moderate impact (i.e. do not reach statistical significance); however, all but two *t* values for the estimates of the interaction between treatment status and FOI* are positive.

Conclusions:

FOI was scaled to range between 0 and 1, therefore the small moderating effect of FOI* (.03 - .04 sd) represents the maximum possible (from zero to full implementation), and is much less than the average effect the study was powered to detect. Power to detect the moderating effect of

FOI* was low. In spite of this, the pattern of differential impacts in Figure 3 is revealing: Adherence matters more to impact than Receptiveness or Participation, and attributes of individuals that, in theory, are related to impact (i.e., based on "Theoretical Model") are shown to actually be so through the Adherence FOI metric but not the others. These results are actionable: FOI* is determined from baseline characteristics and is mildly predictive of impact, which suggests we can use the index to intervene and strengthen the program in cases where we may expect it to be diminished – in this way research can influence practice. (Though beyond the scope of this abstract, in the full paper we discuss how the qualitative data from which Adherence-related fidelity scores were derived suggest places where the program can be improved to boost meaningful participation.) For reference, we include in Table 3 profile descriptions for levels of Adherence on which the quantitative index was based.

References:

Hulleman, Chris S. and Cordray, David S.(2009) Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*, 88–110.

Peck, L. R. (2003). Subgroup analysis in social experiments: Measuring program impacts based on post-treatment choice. *American Journal of Evaluation*, *2*, 157-187.

Peck, L. R. (2016). On the "how" of social experiments: Analytic strategies for getting inside the black box. In L. R. Peck (Ed.), *Social experiments in practice: The what, why, when, where, and how of experimental design & analysis. New Directions for Evaluation*, *152*, 85–96.

Unlu, F., Bozzi, L., Layzer, C., Smith, A., Price, C., & Hurtig, R. (2010, October 13). Linking implementation fidelity to impacts in an RCT: A matching approach. Paper to be presented in symposium: Using Matching Methods to Analyze RCT Impacts on Program-Related Subgroups, Association for Public Policy Analysis and Management. Retrieved from http://www.aefpweb.org/sites/default/files/webform/Uhlu_InPress.pdf

Appendix

Table 1. Alternative measures and metrics for Fidelity of Implementation.

Measure of FOI	Formulas (Metric)	
Adherence (AVERAGE)	Adhr(AVG) = A1 + A2 + A3 + A4	
Adherence (HOLISTIC)	Adhr(HOL)	
	Compensatory	Multiplicative (Conjunctive)
Participation	PC=P1 + P2 + P3 + P4	PM=P1 × P2 × P3 × P4
Receptiveness & Participation	RPC=(P1 + P2 + P3 + P4) + (R1 + R2 + R3)	RPM=(P1 + P2 + P3 + P4) × (R1 + R2 + R3)

Note:

P1="Participation 1": Days attended foundational training
P2=" Participation 2": Days attended monthly synchronous meetings
P3=" Participation 3": Days attended monthly Professional Learning Communities
P4=" Participation 4": Completion of monthly asynchronous assignments
PC= "Participation" metric using Compensatory approach
PM= "Participation" metric using Multiplicative approach
RPC= "Receptiveness and Participation" metric using Compensatory approach
RPM= "Receptiveness and Participation" metric using Multiplicative approach
R1="Receptiveness 1": Sense of level of preparation from training on various dimensions
R2="Receptiveness 2": Perceived levels of helpfulness of the monthly meetings
R3="Receptiveness 3": Perceived levels of helpfulness of the monthly Professional Learning Communities.
A1="Adherence 1": level of use of the core practices of the program
A2="Adherence 2": level of teacher's attention to student thinking.
A3 ="Adherence 3": level of teacher's persistence in problem solving as related to the program
A4 ="Adherence 4": level of teachers' use of a variety of text in ways consistent with the principles of the program
Adh(AVG)="Adherence Average"
Adh(HOL)="Adherence Holistic"
A1 was given 2.5 times the weight of each of A2 – A4; This was decided by the program developer in consultation with the evaluator, and was part of planning for NEi3 reporting.
*When calculating FOI scores, the total Participation score was weighted the same as the total Receptiveness Score.

Table 2. Summary Table Comparing Models used to Obtain Predicted Implementation Values and Differential Impact Estimates

Implementation Metric		R-squared (Proportion of Variance in FOI* explained by teacher-level baseline covariates)	Correlation between FOI (i.e., measured) and FOI* (i.e., predicted) in the treatment condition	Differential impact across levels of predicted implementation
COLUMN:				
PC	Model A	.807 J=35	.903 (p<.001), J=35	-0.018 (SE=.041), DF=1396, t=-.45 p=.653, J=68, J(T)=35, J(C)=33 n=1462,
	Model B	.843, J=33	.918 (p<.001), J=33	0.022 (SE=.029), DF=1348, t=.76 p=.450, J=66, J(T)=33, J(C)=33, n=1414
	Theoretical Model	.416, J=34	.645 (p<.001), J=34	-0.009 (SE=.047), DF=1368, t=-.20 p=.844, J=67, J(T)=34, J(C)=33, n=1434
PM	Model A	.844, J=34	.919 (p<.0001), J=34	0.024 (SE=.030), DF=1368,t=.78 p=.435, J=67, J(T)=34, J(C)=33 n=1434,
	Model B	.758, J=34	.872 (p<.0001), J=34	0.015 (SE=.026), DF=1396,.78 p=.563, J=68, J(T)=34, J(C)=34 n=1462,
	Theoretical Model	.384, J=34	.620 (p<.0001), J=34	0.032 (SE=.040), DF=1368, t=.81 p=.420, J=67, J(T)=34, J(C)=33 n=1434,
RPC	Model A	.844, J=34	.919 (p<.0001), J=34	0.024 (SE=.030), DF=1368, t=.78 p=.435, J=68, J(T)=34, J(C)=34 n=1462,
	Model B	.758, J=34	.872 (p<.0001), J=34	0.015 (SE=.027), DF=1368, t=.58 p=.563, J=68, J(T)=34, J(C)=34 n=1462,
	Theoretical Model	.384, J=34	.620 (p<.0001), J=34	0.032 (SE=.040), DF=1368, t=.81 p=.420, J=67, J(T)=34, J(C)=33 n=1434,
RPM	Model A	.912, J=34	.955 (p<.0001), J=34	0.016 (SE=.036), DF=1396, t=.45 p=.656, J=68, J(T)=34, J(C)=34 n=1462,
	Model B	.685, J=34	.838 (p<.0001), J=34	0.017 (SE=.049), DF=1396, t=.34 p=.733, J=68, J(T)=34, J(C)=34 n=1462,
	Theoretical Model	.425, J=34	.652 (p<.0001), J=34	0.070 (SE=.057), DF=1396, t=1.24 p=.214, J=68, J(T)=34, J(C)=34 n=1434,
Adh (Avg)	Model A	.93 J=29	.962 (p<.001), J=29	.04 (SE=.03) DF=1396, t=1.37 p=.17, J=68, J(T)=34, J(C)=34 n=1462
	Model B	.98 J=29	.988 (p<.001), J=28	.04 (SE=.03) DF=1348, t=1.65 p=.098 J=66,

				J(T)=33, J(C)=33 n=1414
	Theoretical Model	.79 J=29	.889 (p<.001), J=29	.05 (SE=.03) DF=1368, t=1.36 p=.17, J=67, J(T)=34, J(C)=33 n=1434
Adh (Hol)	Model A	.92 J=29	.957 (p<.001), J=29	.03 (SE=.015) DF=1368, t=1.68 p=.094, J=67, J(T)=34, J(C)=33 n=1434
	Model B	.97 J=29	.967 (p<.001), J=29	.02 (SE=.01) DF=1396, t=1.58 p=.114, J=68, J(T)=34, J(C)=34 n=1462
	Theoretical Model	.76 J=29	.872 (p<.001), J=29	.03 (SE=.02) DF=1368, t=1.39 p=.16, J=67, J(T)=34, J(C)=33 n=1434

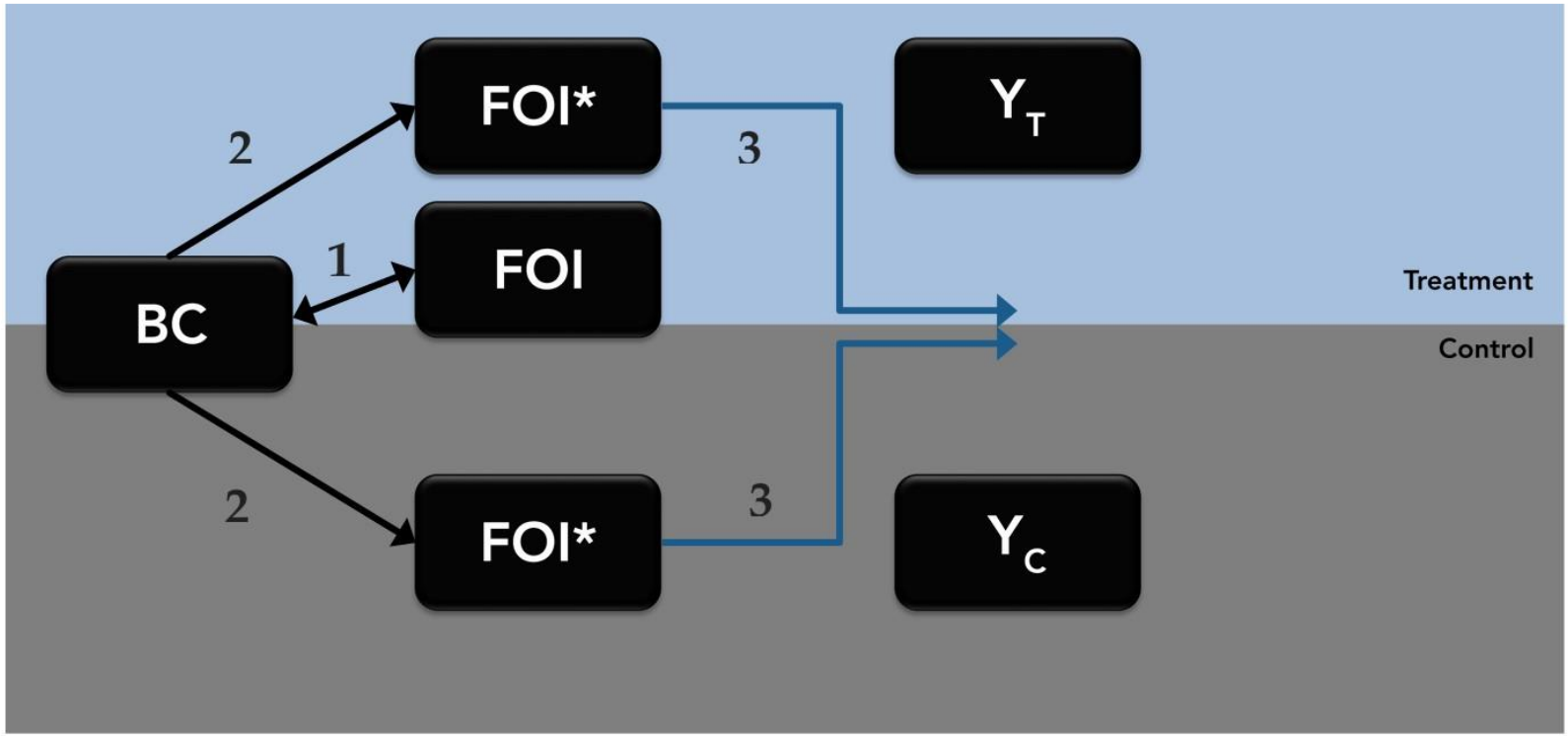
Table 3. Summary of Facilitators' Ratings of Teachers' Levels of Implementation

	0 (some of these also included no evidence.)	1	2	3
Adherence 1: Using core RA practices	Shows no evidence. Prioritizes competing initiatives. Experiences difficulty "meshing methods" using iRAISE and other approaches. Aims to make iRAISE "fit her style".	Struggles with some elements of iRAISE. Is inconsistent in use of specific elements of iRAISE. Tries just some elements. Has difficult getting students to engage.	Tries some core routines. Shows consistent use of core routines. Shows intermittent and experiences challenges but demonstrates effort.	Tries many, including core, routines. Demonstrates multiple iRAISE practices through student work. Shows awareness of the learning culture.
Adherence 2: Attention to student thinking	Shows no evidence. Is in a situation where only some students are engaging. Focuses on 'compliance' over engagement with text, thinking and learning. Focuses on "knowledge attainment"	Focuses on covering content (and related routines, such as teaching students to take detailed notes.) Is frustrated by students' inability to engage in iRAISE strategies. Displays improvement with students' engagement with text. Assumes students already know how to handle difficult text. Considers students to be stalled in depth of thinking. Understands the principles but does not apply them.	Uses evidence of students' thinking to determine implementation success/ challenge. Is concerned that with use of new text students' level of thinking is going down, but is persevering. Uses strategies that continue to evolve and to deepen student thinking. Experiences insight into the process: "today I got that talking to the text is truly the beginning of the conversation". Has insight that by making thoughts visible a student can see misconceptions. Continues to focus on "right/wrong" of content answers. Focuses on students' sense of security in participating in the classroom, but consider it "a work in progress".	Puts student thinking at the center of teaching. Is responsive to student thinking. Student work evidences increased engagement with text over time. Shows student questioning, making connections in student work. increasing engagement with text over time, making inferences, summarizing text. Teacher is "learner-focused and her thoughts center on student engagement and thinking". Expresses concern with student risk-taking in sharing thinking. Stresses the importance of student voice in the class. Recognizes stages students go through in acquiring skills through iRAISE.
Adherence: 3 Persistence in problem-solving implementation	Struggles with "opportunities for reading, student engagement, and student discourse". Focuses on the struggles of implementation. Expresses attitude that additional text functions only to make the subject matter more interesting.	Is receptive to new ideas but shows limited implementation. Admits struggling as a reader herself, Admits lack of confidence in modeling for students. Persists with some elements of iRAISE but gives up on others after trying once. Feels students don't have ability to engage with text on deep level. Does not try to deepen or expand implementation. Believes	Feels supported by the Professional Learning Community, but not by the school. Does not implement deeply or get at core practices. Expresses a lack of confidence in use of methods and using a tinkering approach. Is consistent with implementation, and is constantly strategizing. Consider strategies to support students with "new text", such as modeling with new text. Grapples with overuse of	Resolves challenges by going back to drawing board. Exhibits confident and thoughtful implementation. Is resilient: tweaks and tries again. "Digs into core routines and layers practices…asks probing questions." Uses student work samples to analyze evidence of student thinking. Wants to help other teachers with her approaches. Is persistent in core practices and grows during the year. Is flexible in implementing iRAISE based on students' needs. Is persistent with implementation even if does not seem

	Focuses on perceived need to have students learn to take detailed notes. Says will get to iRAISE strategies after teaching them the basics. Is content to do iRAISE "once in a while".	it is not his/her role to support students with reading in science. Shares that s/he struggled as a learner and is not sure how to support students. Wants more subject-specific examples.	underlining as a strategy by students. Notes time for implementing iRAISE is a challenge. Expresses confidence in making mistakes. States a need to have the opportunity to dialogue with someone in the district. Expresses difficulty finding activity to support text.	to be working. Admits sometimes not knowing "where to go" with student work and responses.
Adherence 4: Use of text	Relies on notes in lecture form. Relies primarily on the textbook for student reading. Is aware of but does not use diverse text.	Uses her notes at text. Says s/he does not have time to find new texts. Tries pulling in additional texts. Relies mostly on the textbook. Exhibits limited use of rich/diverse text.	Tries to increase level of text complexity, but it is hard to find in Chemistry. Uses a range of texts. States challenge of finding more resources.	Uses multiple text types, and increases complexity with time (evident in samples). Grapples with challenges of using text. Describes implications for instruction and learning of using different text / resources.

	0	1	2	3	4	5
Adherence (Holistic)	Focuses on covering content and struggles with social dimension of her class. Gave up in face of "students' apathy". Sees iRAISE as separate entity – students should be "working on it on their own". Shows "minimal perseverance with framework." Implements "at surface". Considers it "hard to get new students on board". Is cautious with thoughtful attention to the few students who are engaging (in an online learning environment); Sees student collaboration as "lacking depth of student thinking and speaking". Expresses concerns with time and grading for completion. Admits "just doing bits for the experience". Does not understand social dimension of RA framework. Responds to student work in a critical way.	Struggles with own reading. Expresses discomfort with certain strategies (metacognitive conversations.) Feels student don't have abilities to do challenging work. Tried some strategies, but struggled with the "social dimension of class". Has limited belief in students' abilities. Focuses on covering content. Struggles with buying into iRAISE. Sees program as an "add on" and students should be "working on it on their own". Discontinues use after a certain amount of time.	Maintains barriers to implementation. Lacks confidence to try new things on her own. Uses some core routines but practice falls flat in terms of deepening student thinking. Shows limited depth in personal and social processes utilizing metacognition, which would contribute to knowledge building. Exhibits less than frequent use of the program.	Looks forward to also using the program the following year. Exhibits exemplary persistence even when struggling with some aspects e.g., building the social/personal dimension in activity.	Supported students to read, think and talk like scientists, but caved in to pressures to cover content from colleagues. Understand iRAISE and relies on colleague for support.	Understands the framework deeply. Becomes program leader on staff. Connects core routines, and personal and social dimensions. Shows evidence of knowledge building that culminates in sense making, and is deeply aware of students' processes: monitoring and documenting their thinking and sharing ideas and knowledge building. Adjusts program strategies based on students' needs.

Figure 1. Steps in the post-experimental method for estimating change of impact with increasing implementation



Steps in method:
(1) Regress FOI against baseline characteristics (BC) in the treatment group,
(2) Use the model to produce predicted values of fidelity (FOI*) for teachers in both conditions,
(3) Assess the interaction between FOI* and the dummy variable for treatment assignment

Figure 2. Assessing adequacy of FOI* prediction model

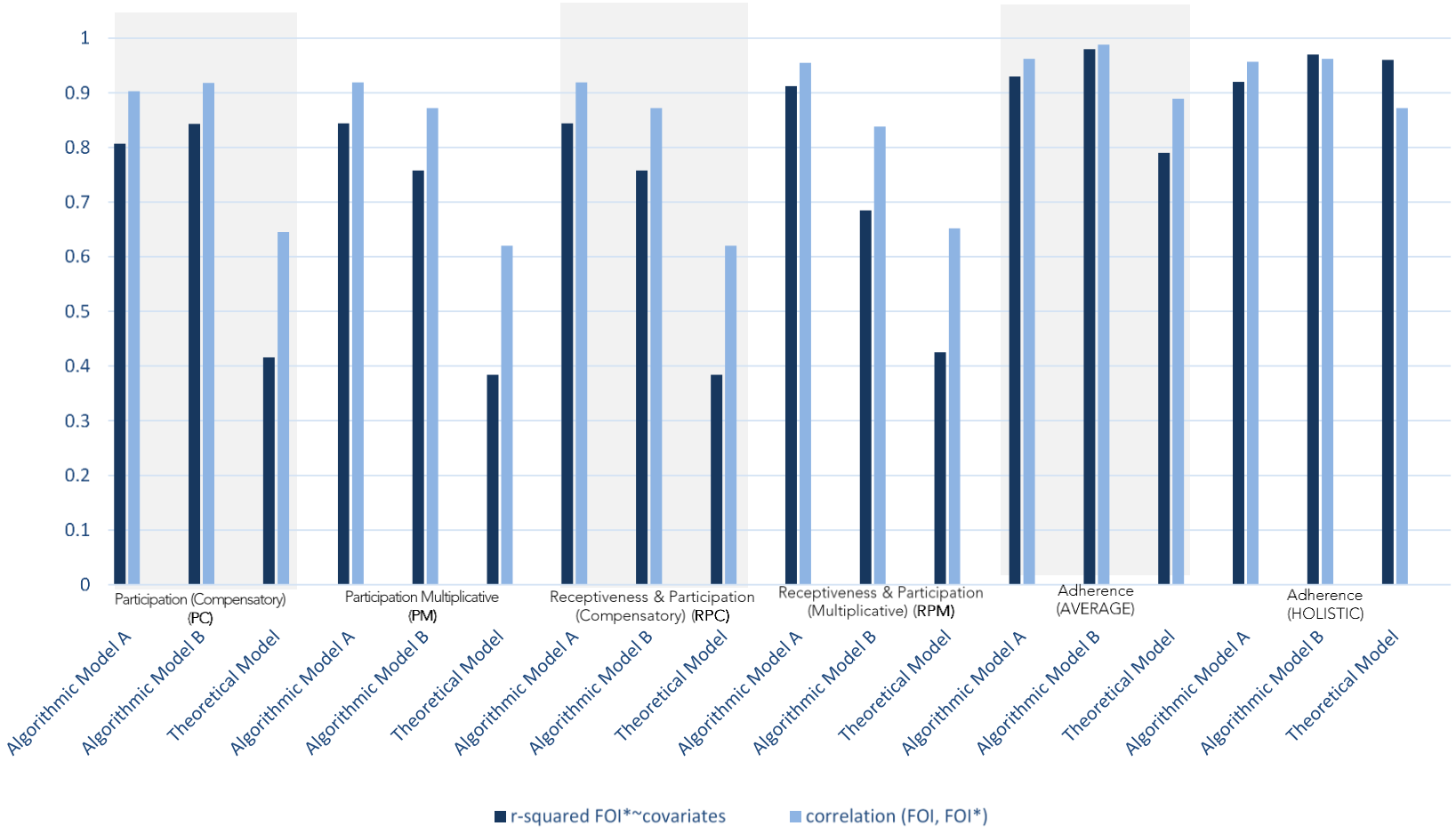
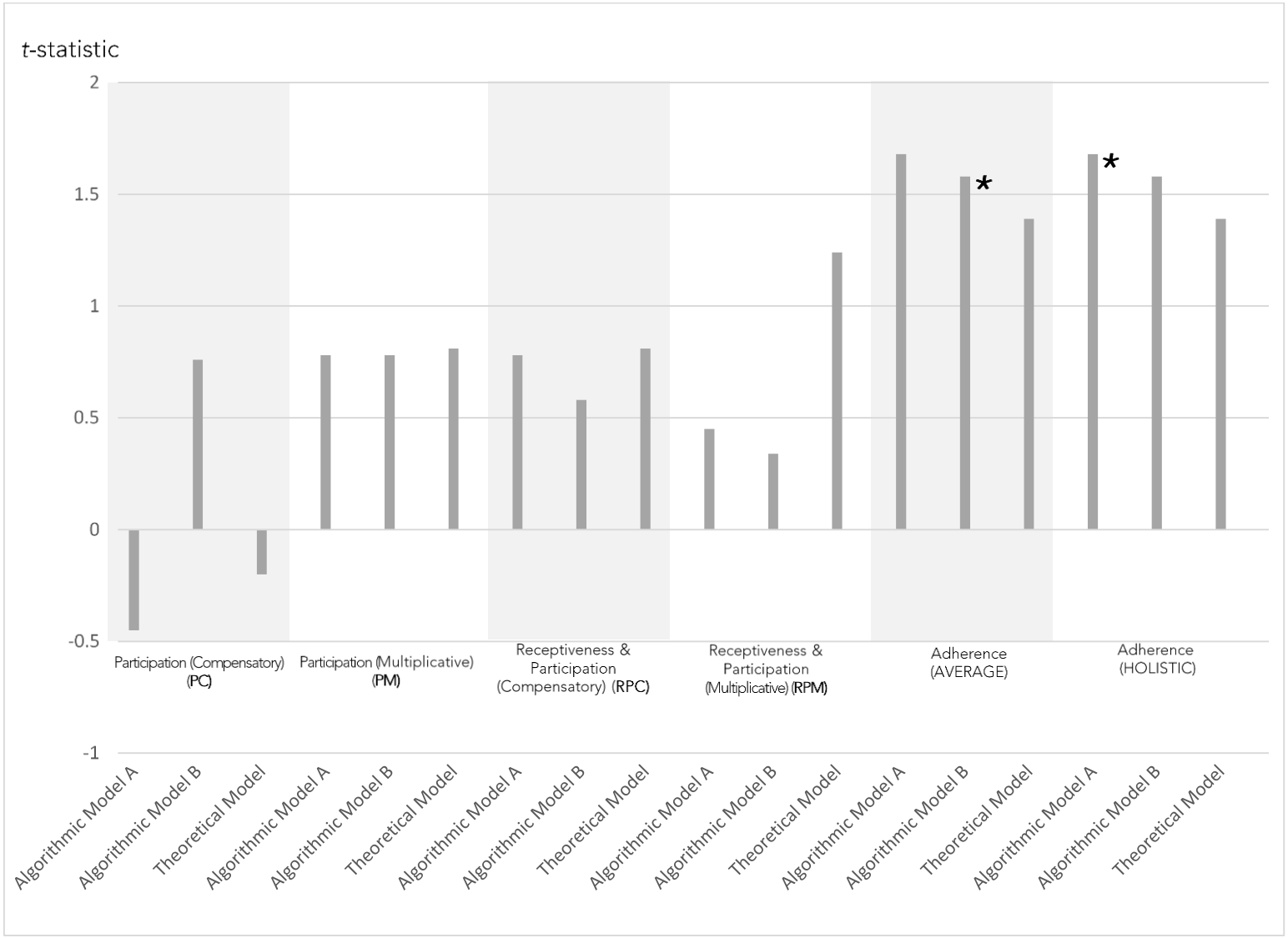


Figure 3. *t*-test statistic for moderating effect of model-predicted fidelity of implementation (FOI*) on impact of iRAISE on student achievement



* $p<.10$