

Swender, E., Surface, E. A., Hamlyn, H., & Wilcox, S. L. (2009, November). *Identifying effective ACTFL oral proficiency testers and raters: Ongoing research*. Symposium presented at the ACTFL 2009 Annual Convention and World Languages Expo, San Diego, CA.

Identifying Effective ACTFL Oral Proficiency Testers and Raters: Ongoing Research

Elvira Swender
ACTFL

Eric A. Surface
SWA Consulting Inc.

Helen Hamlyn
ACTFL

Shelia L. Wilcox
SWA Consulting Inc.



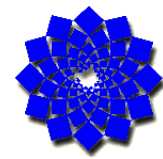
NOVEMBER 2009

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED

Copyright Notice

This document and its content is copyright ©1997-2010 of SWA Consulting Inc. All rights reserved.

Any redistribution or reproduction of part, or the entire document in any form is prohibited except for: (1) you may print or download to a local hard disk extracts for your personal and non-commercial use only, and (2) you may copy the content to individual third parties for their personal use, but only if you acknowledge the website and author(s) as the source of the material. You may not, except with our express written permission, distribute or commercially exploit the content, nor may you transmit it or store it on any other website or other form of electronic retrieval system.

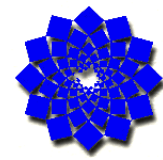


Identifying Effective ACTFL Oral Proficiency Interview (OPI) Testers and Raters: Preliminary Findings from the ACTFL OPI Training Workshops

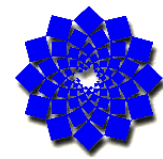
November 21, 2009
ACTFL Annual Convention 2009

Dr. Elvira Swender
ACTFL

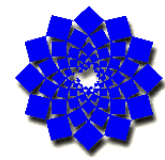
Dr. Eric A. Surface
Ms. Shelia Wilcox
SWA Consulting Inc.



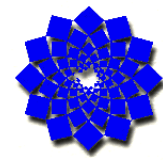
- History and Background of OPI Training
- Rationale for Current Studies
- Research Background & Objectives
- Study Methodology
 - Berkeley through College Park
 - Orlando
- Findings
 - Applying for certification
 - Completing certification
- Conclusions
- Future Directions
- Questions and Answers



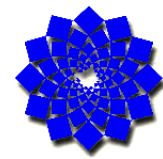
History and Background of OPI Training



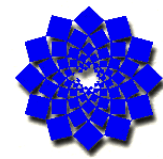
- Berkeley (2004) through College Park (2005)
 - 4-day workshop
 - Each morning begins with a Plenary Presentation
 - Day 1: Focus on ACTFL Scale
 - Day 2: Focus on the Assessment Criteria
 - Day 3: Focus on Interview Structure & Elicitation
 - Day 4: Focus on Implications & Applications of the OPI
 - Each afternoon in language-specific break out groups



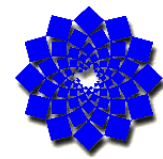
- Day 1
 - Observe and critique a live demonstration interview
 - Listen and Rate sample interviews across levels
 - Begin conducting practice interviews !!!
- Days 2-4
 - Practice interviews conducted by workshop participants
 - Progressing from lower to higher levels
 - Participants rate samples and critique interviews under the guidance of trainer



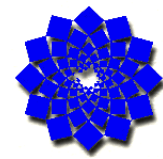
- Optional Certification Process:
 - Apply for OPI Tester Certification
 - Schedule a personal OPI to verify own level of speaking proficiency
 - Conduct *Practice Round* interviews & rate
 - Submit *Practice Round* set of tapes
 - Receive feedback from trainer
 - Submit *Certification Round* set of tapes
 - Recommendation for Certification



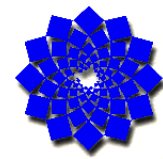
- Presented full rationale at Orlando (2008)
- Highlights:
 - “3rd Day Effect”
 - Skill Interference: Interviewing and Rating
 - Interviewing & Rating are *very* different skills
 - Certification Bottleneck
 - Led to recommendations for improving training
 - “two-tier” training and certification model
 - Rating first, Interviewing second



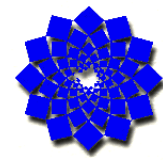
- 4-day workshop
- Currently offered in 8 languages
 - Arabic, Chinese, English, French, German, Japanese, Russian, Spanish
- Each morning begins with a Plenary Presentation
- Each afternoon in language-specific break out groups



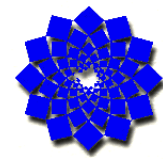
- Days 1 and 2 focus on learning to listen and rate
- Variety of standardized activities
 - Determining at level v. below level performance
 - Differentiating between Mid and Low sublevels
 - Differentiating between sustained and unsustained performance
- Days 3 and 4 focus on elicitation and rating
 - Live practice interviews
 - Trainer and peer critique



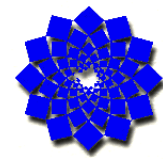
- Submit application
- Personal OPI
- Online Listen and rate Practice Round
- Online Listen and rate Certification Round
 - Must agree exactly on 50% of samples and within a contiguous sub-level on the remaining samples
- Online Rater Support
 - Online listen and rate activities with rationales
 - Online Rater and Tester Tutorials



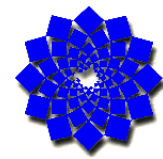
- Must first be a certified OPI Rater
 - Submit application
 - Conduct Phase I interviews and receive feedback**
 - Conduct Phase II interviews
 - Certification recommended
- ** If criteria are met following Phase I, may be recommended for certification at that time



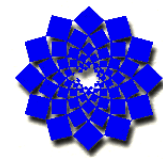
Rationale for Current Studies



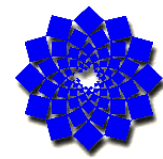
- ACTFL is interested in continuous improvement of its training and certification process
 - Information is required about
 - *For Whom is the training Effective?*
 - *Who should apply for certification?*
- Impact on training design
 - *How can training be optimized for different learners?*
- Impact on recruiting successful trainees
 - *Critical for meeting the demand for certified testers and raters in the profession*
- Impact on selecting successful trainees for other programs, such as ILR Tester training



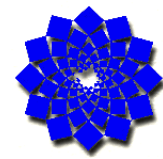
Research Background and Objectives



- The effectiveness of any rater-based assessment depends on the extent to which all the raters share a standardized ***mental model*** of proficiency and a standardized protocol for eliciting and rating that proficiency and apply them consistently.
 - ***Having an effective training and certification process is paramount to having an effective assessment.***
 - ***Presentation in Orlando (Surface, Swender, Brown & Dierdorff, 2008) demonstrated the effectiveness of the training process and the change in training format.***

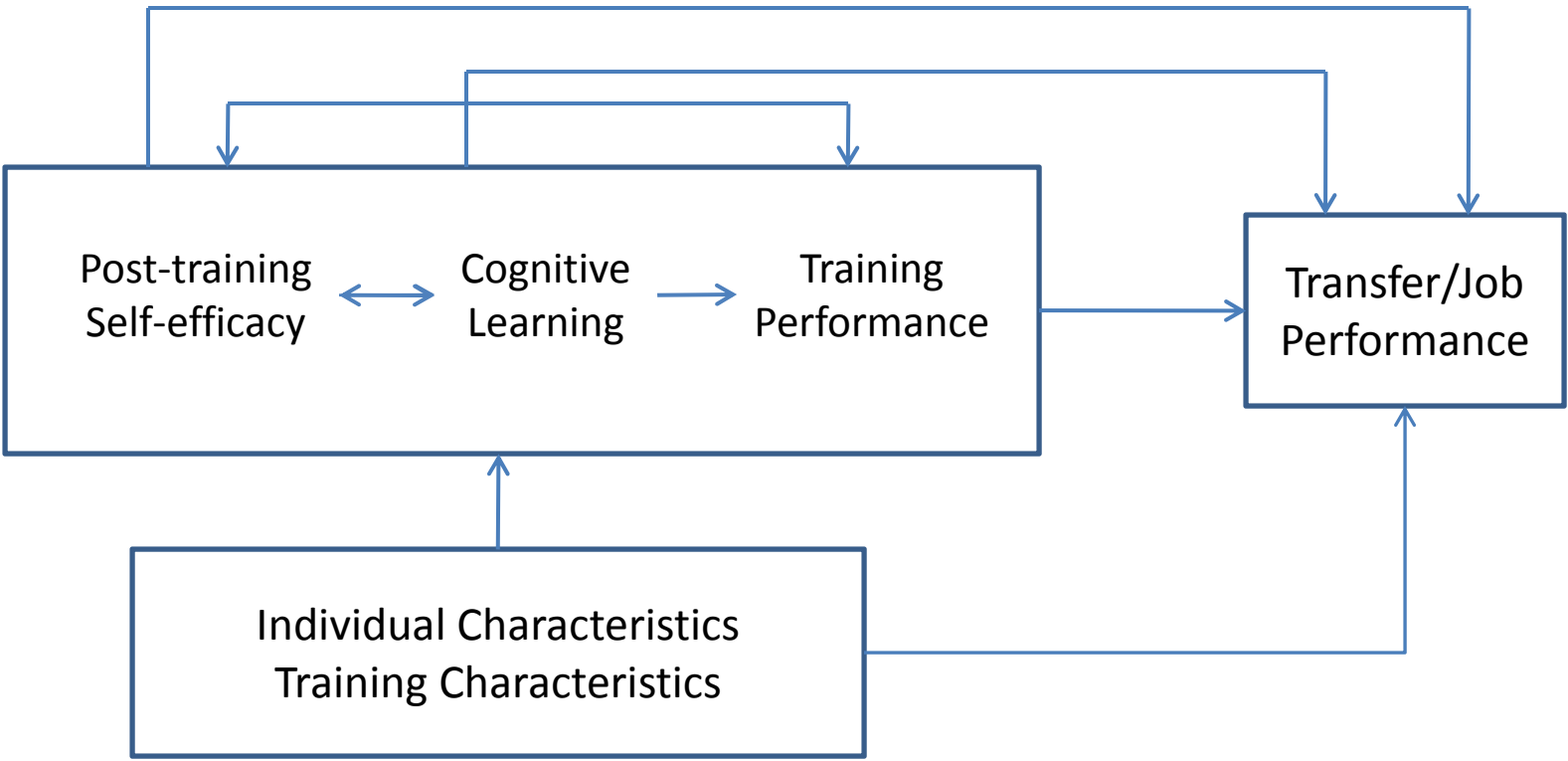
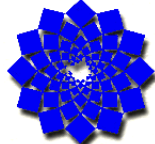


- To become a successful OPI tester or rater for ACTFL, there is a gated process.
- Multiple hurdles must be navigated:
 - Training
 - Application for certification
 - Certification
 - Working as a tester
- Who are the people who best navigate the process?

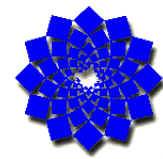


- Our ultimate criterion would successful job performance as an ACTFL-certified Tester or Rater
 - Sufficient performance data are not available at this time
 - *Since certification is required for performance, we are focusing on certification outcomes*
 - *Training Outcomes—deferred for time*
 - *Publication is under review*
- What factors are related to certification outcomes?
 - What types of people apply and become certified?
 - What factors predict:
 - *Applying for certification?*
 - *Obtaining certification?*

Research Framework

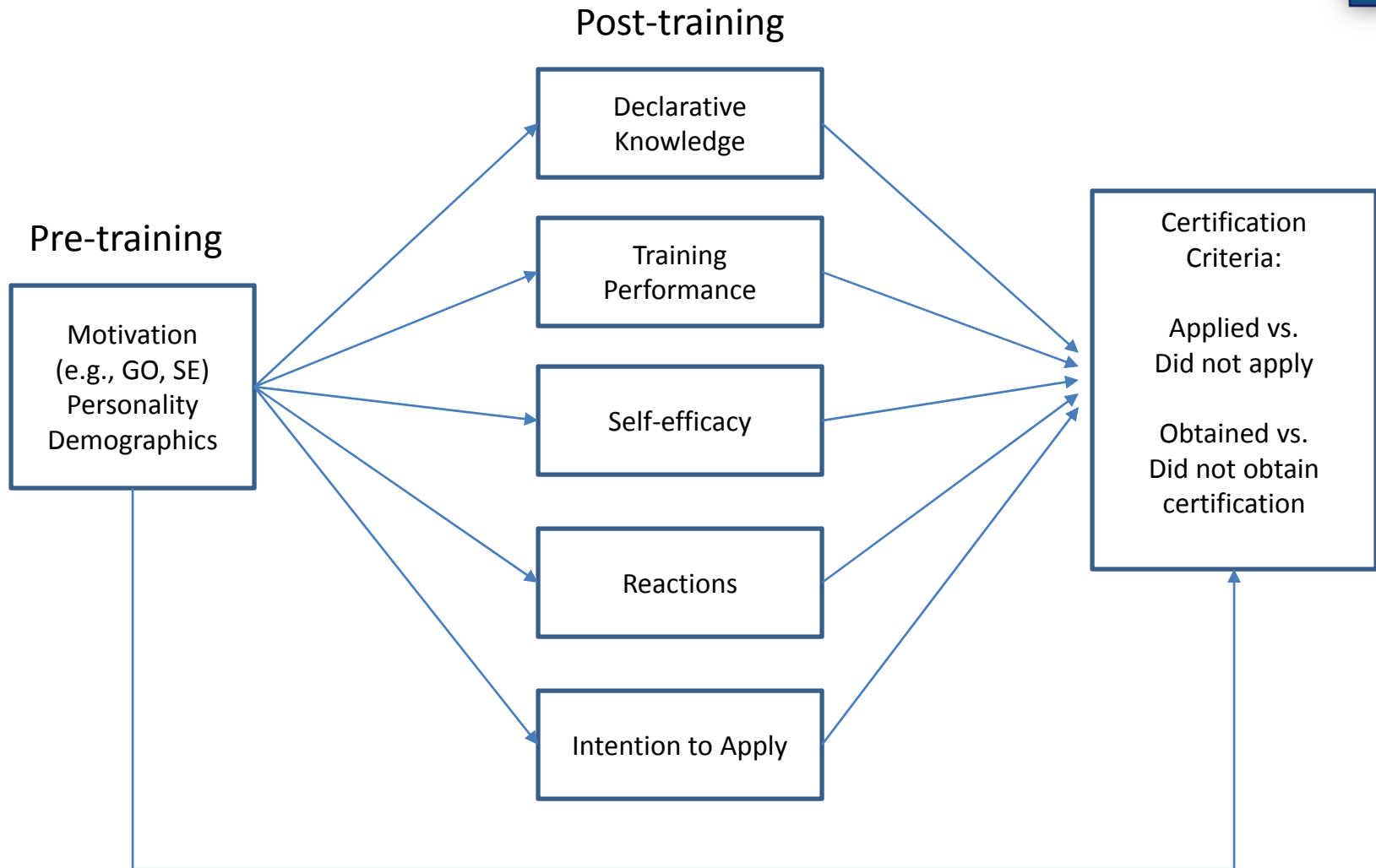
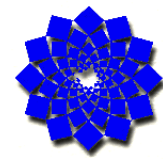


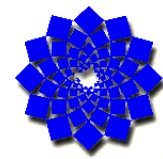
Adapted from Alvarez, K., Salas, E., & Garafano, C.M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review*, 3(4), 385-416.



- Individual differences
 - Motivation, Personality, Demographics
- Training Outcomes
 - Post-training Declarative Knowledge
 - Training Performance
 - Post-training Self-efficacy
 - Workshop Reactions
 - Post-training Intentions
- *The goal of training is to standardize testers and raters on these training outcomes criteria, so there may be no variability (i.e., No prediction)*

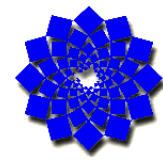
Tested Model



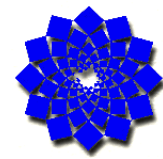


Study 1 Methodology

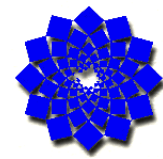
Berkeley through College Park



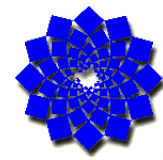
- Conducted at 4 Locations:
 - Berkeley (Pilot; July 2004)
 - Chicago (November 2004)
 - Utah (July 2005)
 - College Park (November 2005)
- Some differences between locations, but mostly same data collected.
 - Changed some after Berkeley



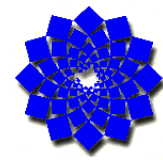
- What SWA collected:
 - Pre-Training Questionnaire
 - Pre-Training Declarative Knowledge Test
 - Participant Proficiency Ratings
 - Post-Training Questionnaire
 - Post-Training Declarative Knowledge Test
- What ACTFL collected:
 - Certification Data
 - Job Performance Data as Testers (in process)



- Training Study: PRE-POST without control group design
- Training Criteria
 - Reactions
 - Learning
 - ABCs (Kraiger, 2002; Kraiger, Ford, & Salas, 1993)
 - Affective Learning
 - Behavioral Learning
 - Cognitive Learning
 - Transfer
 - Certification
 - Intentions to apply for certification
 - Application for certification
 - Certification Results

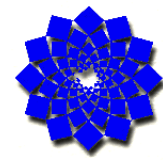


- Pre-Training Questionnaire
 - $N = 254$
 - Berkeley = 31, Chicago = 92, Utah = 102, College Park = 29
 - Demographic Items
 - Task-Specific Self-Efficacy
 - Intentions
 - Psychological
 - Goal Orientation, Learning Self-Efficacy, Valence of Training, Personality
- Pre-Training Declarative Knowledge Test
 - Not included in the Berkeley Study
 - $N = 223$



- Post-Training Questionnaire
 - $N = 301$
 - Berkeley = 62, Chicago = 95, Utah = 32, College Park = 112
 - Reactions
 - Satisfaction, Effectiveness, Open-ended comments
 - Task-Specific Self-Efficacy
 - Intentions
- Post-Training Declarative Knowledge Test
 - $N=239$
 - Match with Pre-Training $N = 202$

Study 1 Training Effectiveness



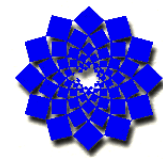
SWA

Consulting

Construct	N	Pre-Training Mean	Post-Training Mean	Change (Post-Pre)
Task-Specific Self-Efficacy ¹	228	4.57	5.90	1.33*
Declarative Knowledge	201	7.56	9.81	2.25*

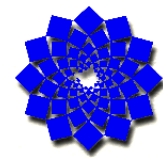
Construct ¹	Post-Training Mean
Perceived Training Effectiveness	6.09
Training Satisfaction	6.04

¹ Scale ranges from 1(lowest) to 7(highest). * $p < .05$.



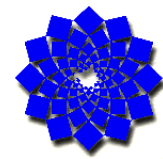
Frequencies		
	Yes	No
Applied for Certification	157 (49%)	163 (51%)
Obtained Certification*	71 (45%)	86 (55%)

* Totals are only for individuals who applied for certification.

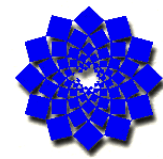


Study 2 Methodology

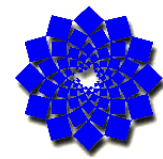
Orlando



- What remained the same?
 - Same number of days of training
 - Same objectives
 - Same content
 - ACTFL Scale
 - Assessment Criteria
 - Interview Structure and Elicitation
 - Implications and Applications of the OPI

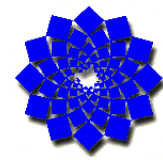


- Changes from Study 1
 - Days 1 and 2 focus on learning to listen and rate
 - Variety of standardized activities
 - Determining at level v. below level performance
 - Differentiating between Mid and Low sublevels
 - Differentiating between sustained and unsustained performance
 - Days 3 and 4 focus on elicitation and rating
 - Practice testing does not begin until Day 3
 - New certification categories and procedures

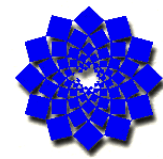


Orlando (2008)

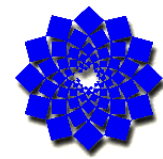
- What SWA collected:
 - Pre-Training Questionnaire
 - Pre-Training Declarative Knowledge Test
 - Daily Evaluations (Day 1, Day 2, & Day 3)
 - Post-Training Questionnaire
 - Post-Training Declarative Knowledge Test
- What ACTFL collected:
 - Certification Data
 - Job Performance Data in Process



- Pre-Training Questionnaire
 - $N = 63$
 - Demographic Items
 - Task-Specific Self-Efficacy
 - Intentions
 - Reason for Attending Workshop
 - Psychological
 - Goal Orientation, Learning Self-Efficacy, & Valence of Training
- Pre-Training Declarative Knowledge Test
 - $N = 62$

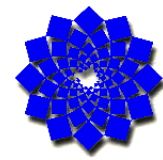


- Participant Demographics:
 - Section Language
 - English = 30, French = 14, Spanish = 39
 - Native Language
 - English = 25, Other = 38
 - Occupation
 - Professor = 43, K-12 Teacher = 6, Other = 12
 - Highest Degree Obtained
 - BA/BS = 6, MA/MS = 22, Ph.D./Ed.D. = 33, Other = 2



- Post-Training Questionnaire
 - $N = 61$
 - Demographic Items
 - Reactions
 - Satisfaction
 - Effectiveness
 - Open-ended Comments
 - Task-Specific Self-Efficacy
 - Intentions
- Post-Training Declarative Knowledge Test
 - $N=59$
 - Match with Pre-Training $N = 53$

Study 2 Training Effectiveness



SWA

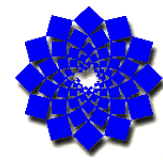
Consulting

Construct	N	Pre-Training Mean	Post-Training Mean	Change (Post-Pre)
Task-Specific Self-Efficacy ¹	58	4.71	6.60	1.89*
Declarative Knowledge	53	8.72	10.91	2.19*

Construct ¹	Post-Training Mean
Perceived Training Effectiveness	6.28
Training Satisfaction	6.21

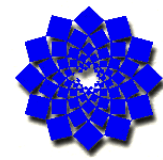
¹ Scale ranges from 1(lowest) to 7(highest). * $p < .05$.

Study 2 Method

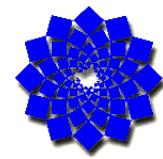


Frequencies		
	Yes	No
Applied for Certification	31 (47%)	35 (53%)
Obtained Certification*	4 (13%)	27 (87%)

* Totals are only for individuals who applied for certification.

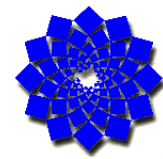


- Path analysis
 - Study 1 (4 locations) samples only
 - Simultaneously assess relationships among all study variables
- Separate logistic regression, linear regression and chi-square analyses
 - Study 2 Orlando sample only
 - Examined prediction of each variable individually
 - Small sample size prohibited more complex analysis



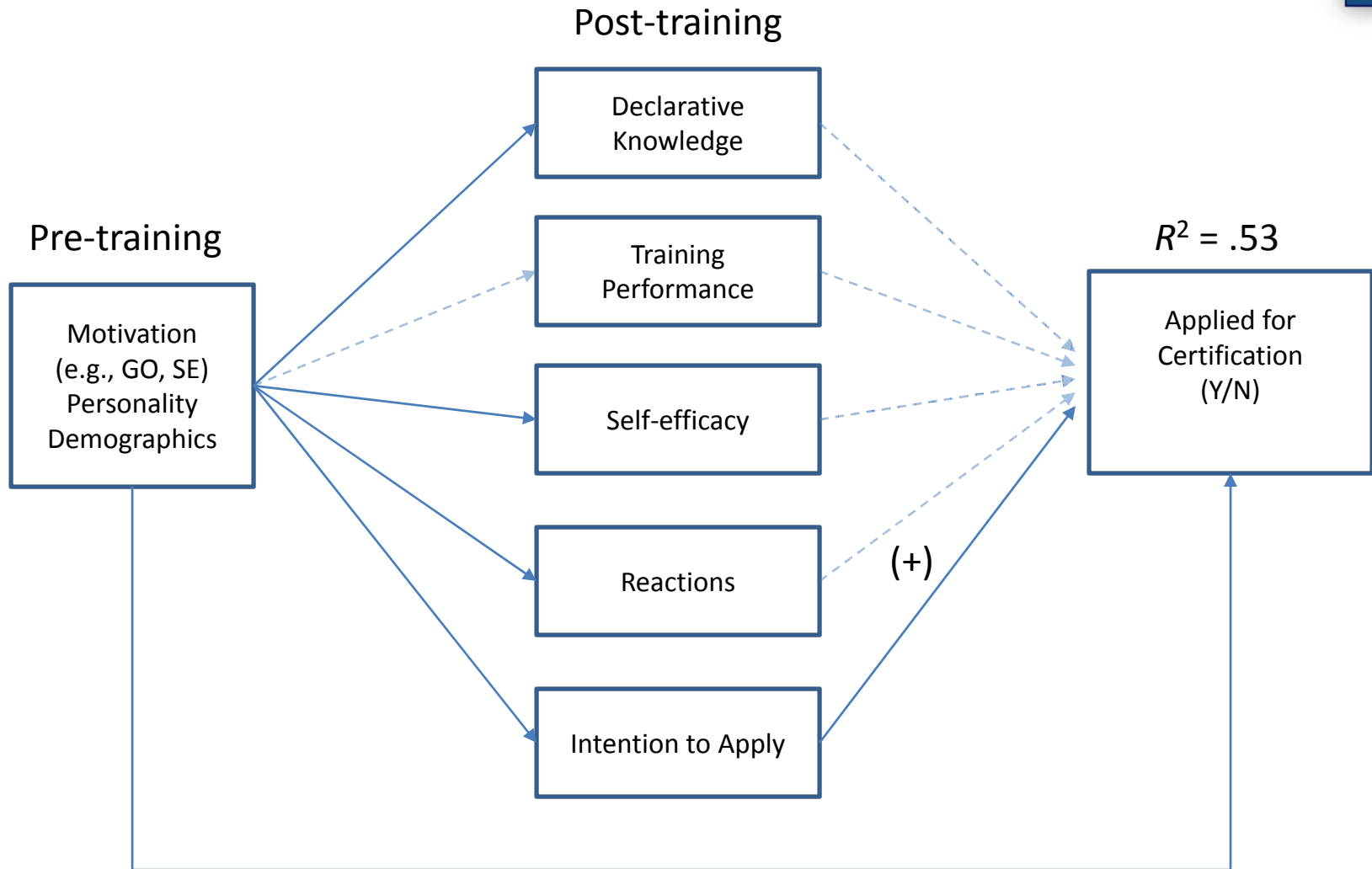
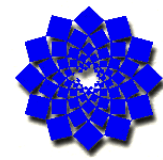
Findings Research Question 1

Who applies for certification?



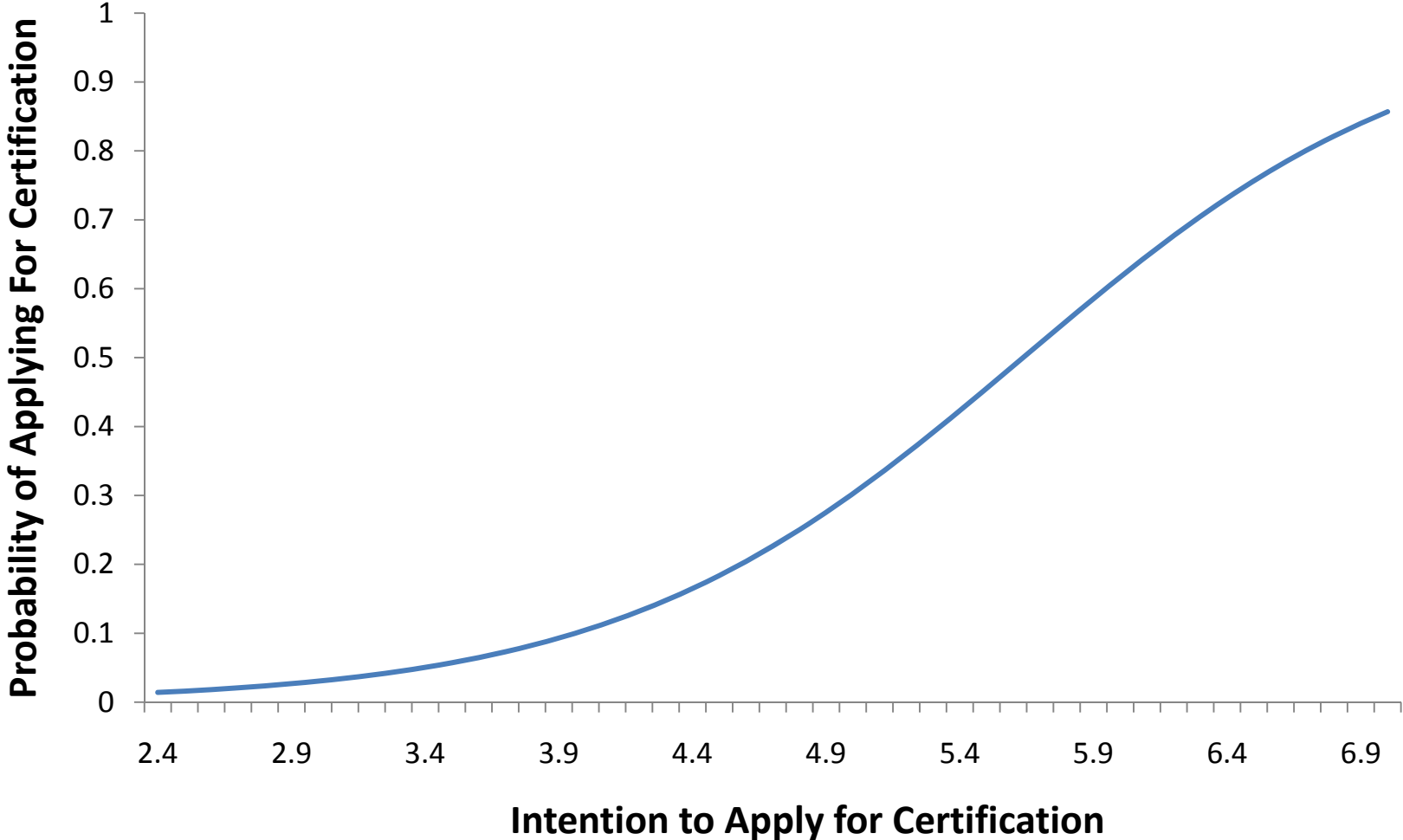
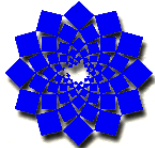
- What factors predict ***applying for certification in Study 1?***
 - *Study 1 Highlights:*
 - *Post-training intention to apply*
 - Greater intention, more likely to apply
 - *Pre-training mastery goal (learning) orientation*
 - Greater mastery orientation, *less* likely to apply

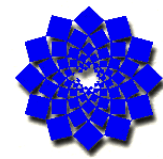
Results for Applying for Certification



Solid lines are significant ($p < .05$). Dotted lines nonsignificant.

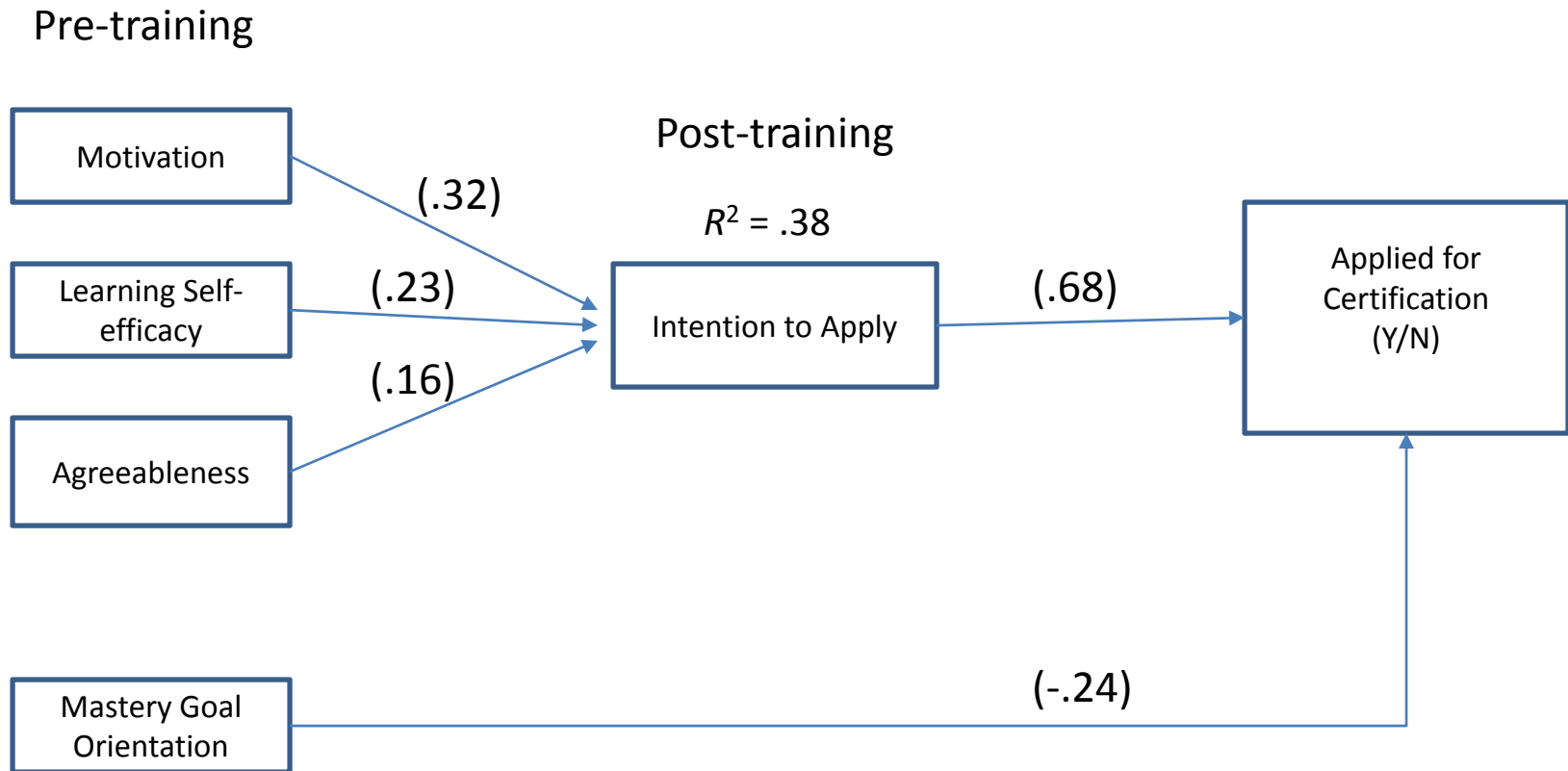
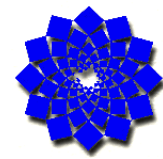
Results for Applying for Certification



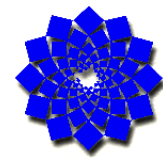


- If *intention to apply* for certification is our single strongest predictor, what determines *intention to apply*?
- What factors predict greater *intention to apply*?
 - *Study 1:*
 - *Pre-training motivation to train*
 - Greater motivation, stronger intention to apply
 - *Pre-training learning self-efficacy*
 - Greater confidence to learn, stronger intention to apply
 - *Agreeableness*
 - More agreeable individuals showed stronger intention to apply

Results for Applying for Certification



Solid lines are significant ($p < .05$). Regression coefficients are standardized. Nonsignificant variables and paths are not shown for clarity.



- What factors predicted ***applying for certification in Study 2?***

- *Study 2 Highlights:*

- *Post-training intention to apply*

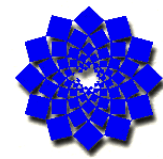
- Greater intention, more likely to apply
- Odds Ratio: 2.474, $p < .01$

- *Pre-training task-specific self-efficacy*

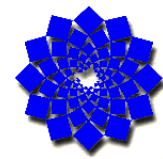
- Greater pre-training task-specific SE, more likely to apply
- Odds Ratio: 1.942, $p < .01$

- *Occupation*

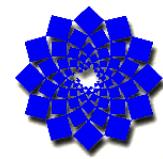
- Attendees who were K-12 teachers were less likely to apply than those in other occupations
- $\chi^2 = 6.44, p = .01$



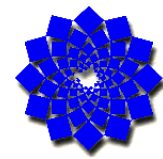
- What factors predict greater *intention to apply in Study 2*?
 - *Study 2 Highlights:*
 - *Performance approach goal orientation*
 - Stronger performance approach, stronger intention to apply
 - $\beta = .259, p = .047$
 - *Task-specific self-efficacy*
 - Higher task-specific SE, stronger intention to apply
 - $\beta = .363, p < .01$



- What factors predict greater *intention to apply in Study 2*?
 - *Study 2 Highlights:*
 - *Reason to become certified*
 - Those whose reason to attend the workshop was to become certified showed greater intention to apply
 - $\beta = .274, p = .036$
 - *Occupation*
 - Individuals who were K-12 teachers showed lower intention to apply for certification
 - $\beta = -.322, p = .014$

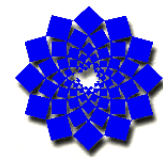


- Summary
 - *Post-training intention to become certified was strongest predictor of applying for certification across both studies*
 - *Motivation and Attitudes individuals brought into the workshop were fairly predictive of who went on to apply*
 - *Learning outcomes of training (e.g., knowledge, skill demonstration) were less predictive*



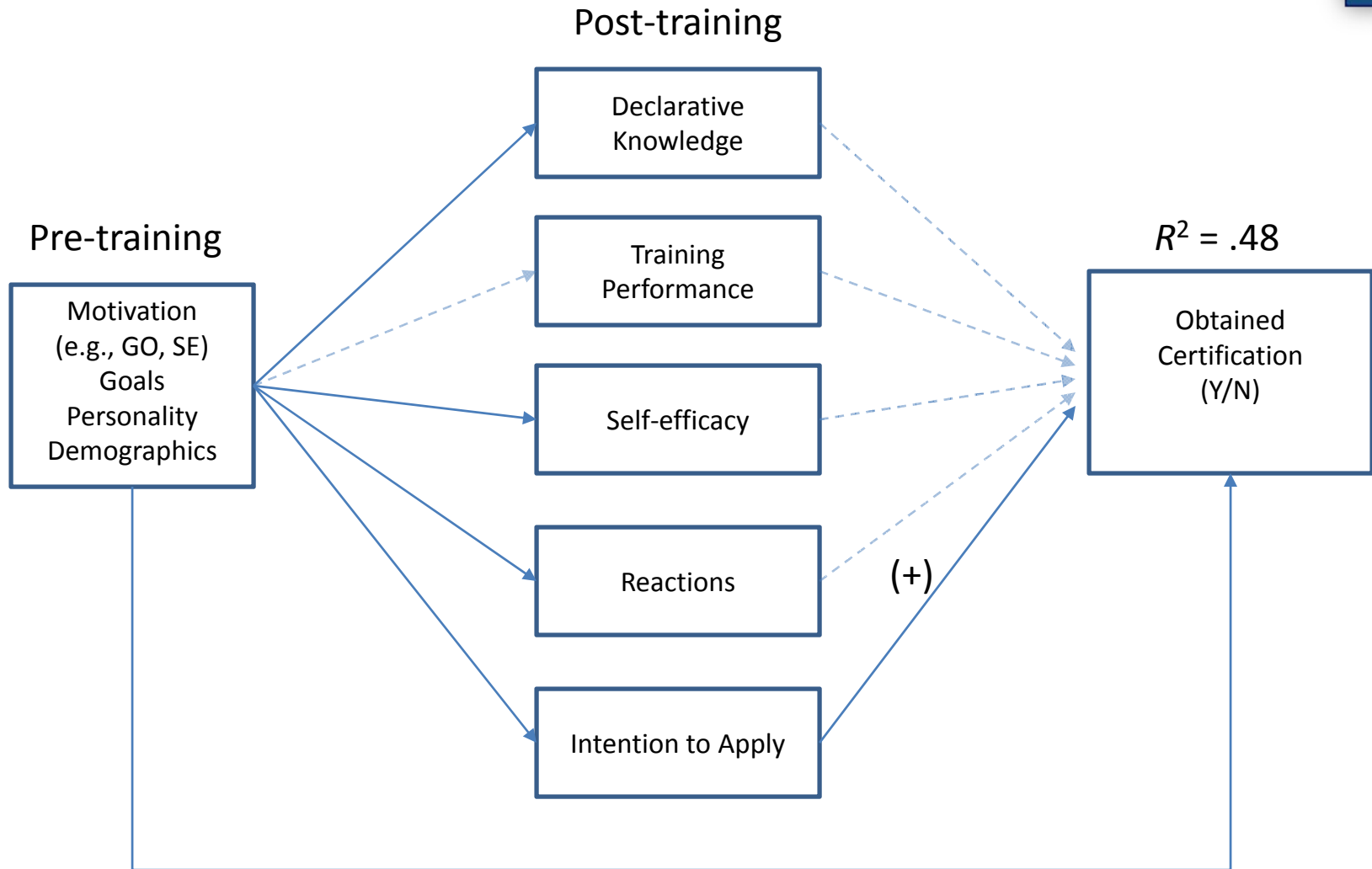
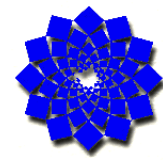
Findings Research Question 2

Who completes certification?



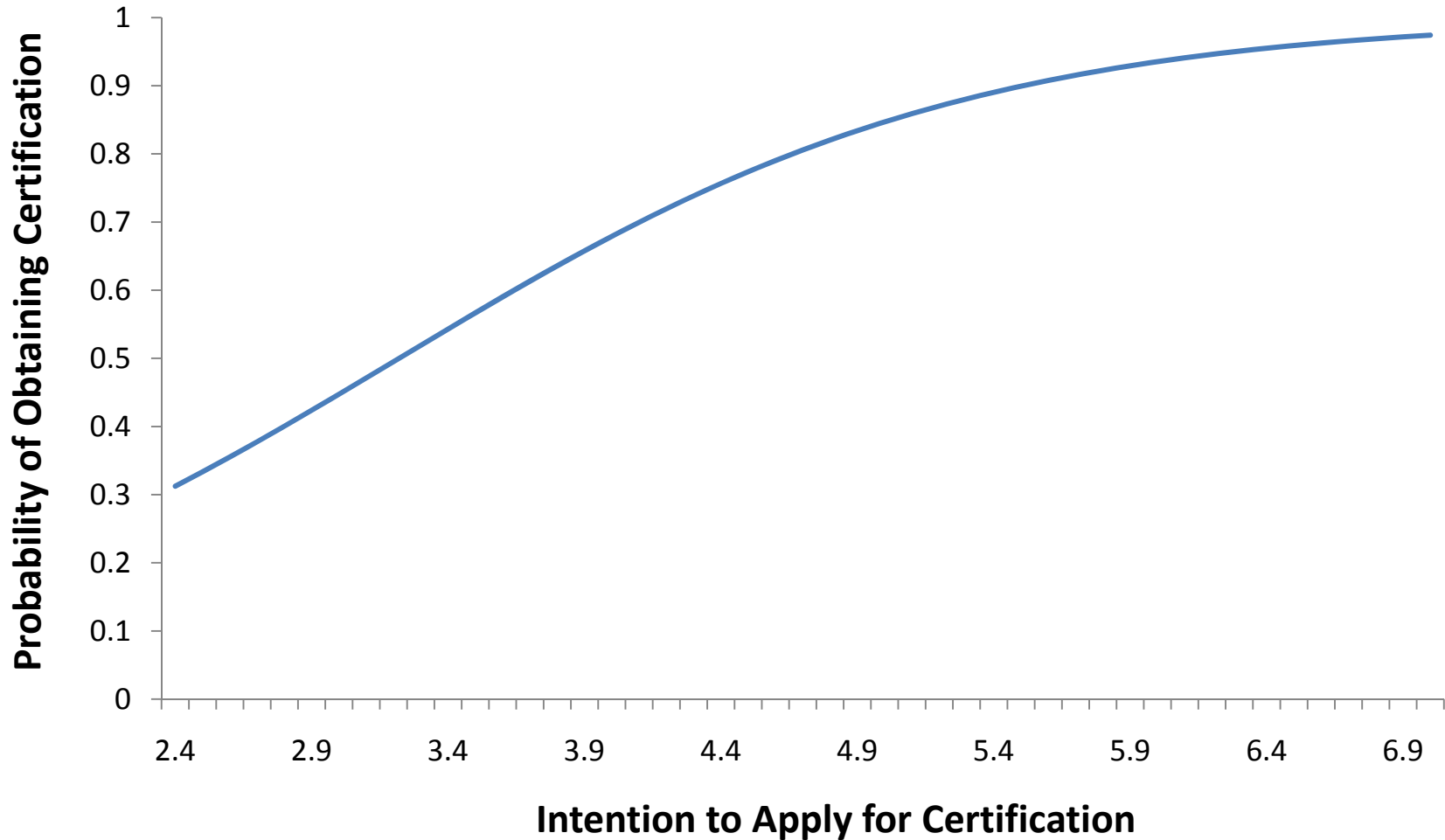
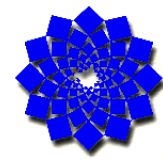
- What factors predict obtaining certification?
 - *Study 1 Highlights:*
 - *Post-training intention to apply for certification*
 - Greater intention, more likely to obtain certification
 - *Pre-training mastery goal orientation*
 - Greater mastery orientation, *less* likely to obtain certification
 - *Pre-training avoidance goal orientation*
 - Greater avoidance orientation, less likely to obtain certification
 - *Emotional Stability*
 - Greater emotional stability, more likely to obtain certification

Results for Obtaining Certification

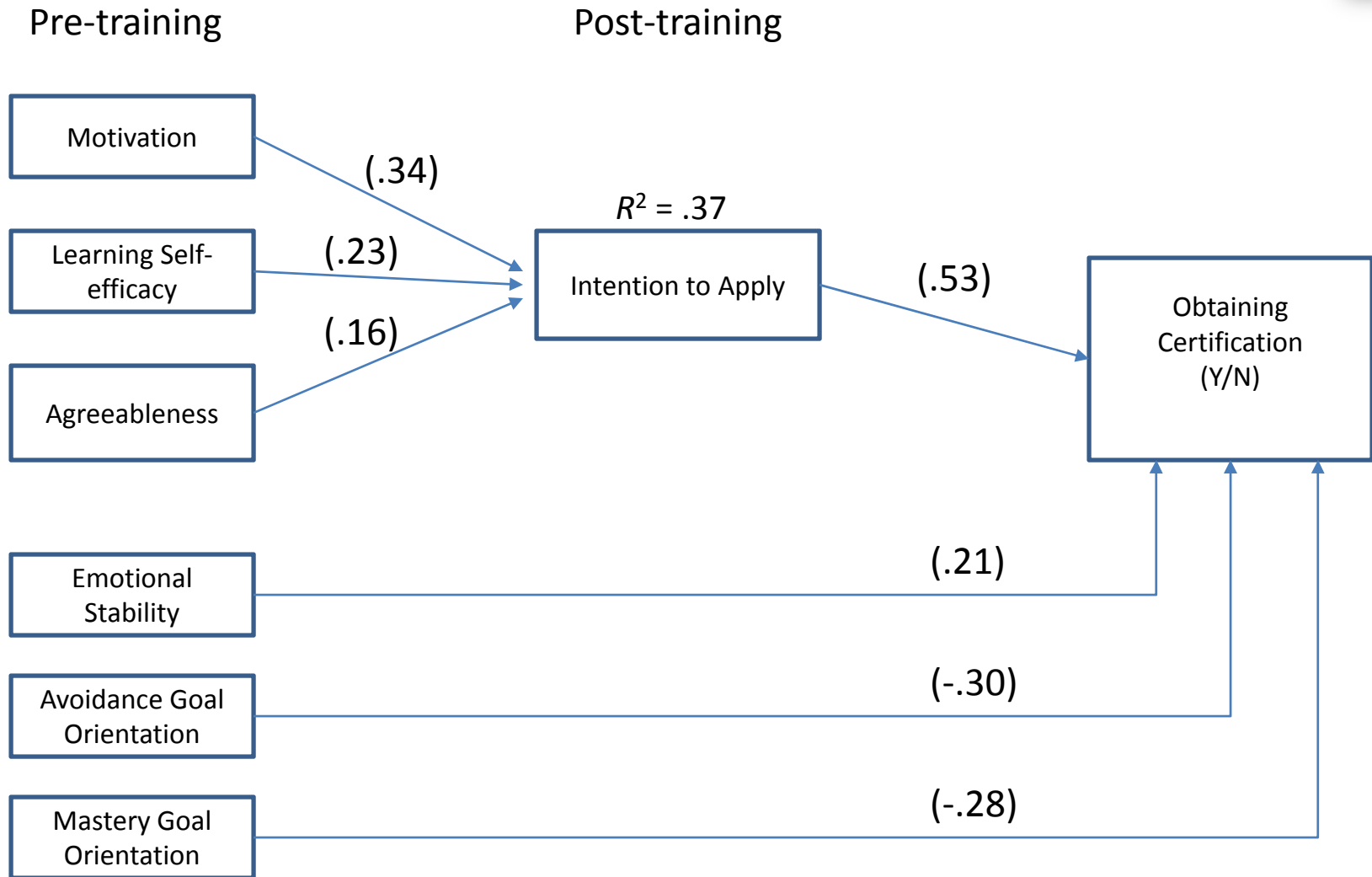
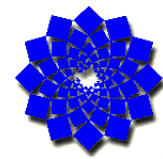


Solid lines are significant ($p < .05$). Dotted lines nonsignificant.

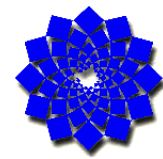
Results for Obtaining Certification



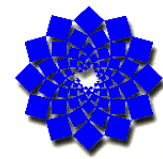
Results for Obtaining Certification



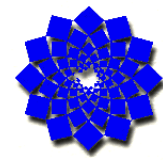
Solid lines are significant ($p < .05$). Regression coefficients are standardized. Nonsignificant variables and paths are not shown for clarity.



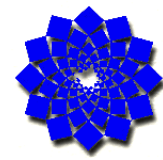
- Summary:
 - *Investigated Study 1 Certifications*
 - *Post-training intention to become certified was the strongest predictor of obtaining certification*
 - *Several dispositional characteristics were found to relate to certification success*
 - E.g., those with the tendency of withdrawing from difficult learning/performance situations (i.e., avoiders) were less likely to obtain certification
 - *Too few individuals in Study 2 obtained certification to include in this analysis*



Conclusions and Future Directions

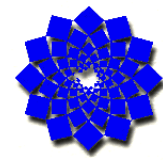


- Implications for certification
 - Motivations and Attitudes (i.e., dispositional characteristics) are important indicators of eventual participation in certification process
 - Data indicate less than half of participants went on to apply for certification
 - Do we do a good enough job promoting certification?
 - Are there barriers that prevent motivated individuals from attending the workshop and/or apply for certification?
 - Do we need to recruit and fund high probability attendees?
 - Data indicate many individuals do not complete the certification process
 - What are the barriers to completing?

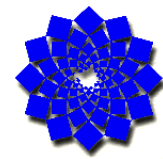


- Implications for training
 - The assumption that most people who attend training want to become a certified testers/rater may be tenuous
 - E.g., Only 44% of respondents (28 of 63) in Study 2 reported this as a reason they attended the workshop
 - Comments suggest many attend to improve their instruction and understanding of proficiency
 - Other vehicles to recruit individual to become certified testers should be considered
 - Evidence did not show linkage between training outcomes and certification success
 - Knowledge and skill acquired in training are necessary but they do differentiate who goes on for certification or who completes*
 - Note: limited data and lack of variability may have impacted results

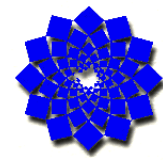
Limitations



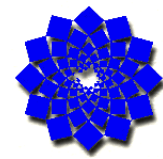
- Training and certification procedures and data collected collected varied across studies, preventing aggregation
- Limited number of testers and raters complete certification
- Job performance were not available at this time
- Training and certification data loss (or lack of complete data) limits sample size impacting analysis options and results
- Need more detailed data from training and certification processes (e.g., interview level data) are needed to actually link training outcomes to certification success



- Finalize the analyses predicting training and certification and publish
- Analyze and present OPIc rater training study data
- Extend analyses to job performance data when sufficient data are available
- Adapt our training and certification study
 - Extend to training not associated with convention
 - More actual training and certification performance data would be beneficial
 - Improve training and certification data collection and management
- Conduct a concurrent validation study
 - Data losses and lack of rater job performance data from the workshop studies prevent us from thoroughly investigating what makes a ‘good’ rater at this time
 - A concurrent validity study could assess qualities embodied by successful tester and raters
 - Uses current testers and raters
 - Validate profile with trainees and participants in the certification process



Closing Thoughts



Questions and Answers

Please feel free to contact
Dr. Swender or Dr. Surface
with additional questions.

ABOUT SWA CONSULTING INC.

SWA Consulting Inc. (formerly Surface, Ward, and Associates) provides analytics and evidence-based solutions for clients using the principles and methods of industrial/organizational (I/O) psychology. Since 1997, SWA has advised and assisted corporate, non-profit and governmental clients on:

- Training and development
- Performance measurement and management
- Organizational effectiveness
- Test development and validation
- Program/training evaluation
- Work/job analysis
- Needs assessment
- Selection system design
- Study and analysis related to human capital issues
- Metric development and data collection
- Advanced data analysis

One specific practice area is analytics, research, and consulting on foreign language and culture in work contexts. In this area, SWA has conducted numerous projects, including language assessment validation and psychometric research; evaluations of language training, training tools, and job aids; language and culture focused needs assessments and job analysis; and advanced analysis of language research data.

Based in Raleigh, NC, and led by Drs. Eric A. Surface and Stephen J. Ward, SWA now employs close to twenty I/O professionals at the masters and PhD levels. SWA professionals are committed to providing clients the best data and analysis with which to make solid data-driven decisions. Taking a scientist-practitioner perspective, SWA professionals conduct model-based, evidence-driven research and consulting to provide the best answers and solutions to enhance our clients' mission and business objectives. SWA has competencies in measurement, data collection, analytics, data modeling, systematic reviews, validation, and evaluation.

For more information about SWA, our projects, and our capabilities, please visit our website (www.swa-consulting.com) or contact Dr. Eric A. Surface (esurface@swa-consulting.com) or Dr. Stephen J. Ward (sward@swa-consulting.com).