

RESEARCH IN PROGRESS REPORT:

Industrial and Organizational Psychology Programme at the University of Colorado at Denver

This article describes the Industrial and Organizational Psychology Programme at the University of Colorado at Denver (CU-Denver). First, we provide an overview of the programme's most salient characteristics and objectives. Next, we summarize five ongoing research programmes being conducted at CU-Denver: (a) Social Power and Influence in Organizations, (b) Personnel Selection, (c) Applied Research Methods, (d) Training Evaluation, and (e) Cognitive Assessment. For each research programme we describe (a) primary questions addressed, (b) selective recent findings, and (c) current projects and future directions. Finally, we describe the role and current activities of the Centre for Applied Psychology

The University of Colorado at Denver offers a Master's degree in Industrial and Organizational (I/O) psychology. The primary objective of the programme is to train students to perform psychological research, evaluation and services in public or private sector organizations. Students receive state-of-the-art training in theories and methods in I/O psychology which can also allow them to pursue further graduate study. The core I/O psychology faculty includes Professors Herman Aguinis and Kurt Kraiger. Students also take electives from the School of Business with professors such as Wayne F. Cascio, Edward O'Connor, Kenneth Bettenhausen and Raymond Zammuto. The curriculum provides a balance between theoretical and applied perspectives. Admission into the programme is very competitive. We admit approximately six to eight students from over 60 applicants each year. Successful applicants usually have at least a 3.5 undergraduate grade point average and 1200 Graduate Record Examination. (GRE) verbal and quantitative scores. Graduation requirements include a total of 32 semester hours, the successful completion of a competency-based comprehensive examination (Aguinis and Kraiger in press) and the successful completion of a thesis or internship (some students elect to complete both a thesis and an internship).

Since the start of the programme in 1983, most graduates have found full-time employment with relatively short search times. Some examples of graduate placements include: assessment specialist for a small I/O psychology consulting firm, personnel researcher for a large telecommunications company and personnel analyst for a city government. Also, some graduates from our programme decided to

further their education and pursue a Ph.D. degree. Recent graduates have been accepted into doctoral I/O psychology programmes in such schools as the University of California at Berkeley and Colorado State University.

Next, we describe five ongoing, programmatic research programmes being conducted at CU-Denver: (a) Social Power and Influence in Organizations, (b) Personnel Selection, (c) Applied Research Methods, (d) Training Evaluation and (e) Cognitive Assessment. These research agendas cover a wide range of topics and speciality areas within I/O psychology.

Research programmes

Social power and influence in organizations

Primary questions addressed. Social power and influence are topics of paramount importance in organizations because they are often regarded as critical determinants of managerial success and organizational advancement (Ragins and Sundstrom 1989). Power is defined as the *ability* or *potential* to alter other people's beliefs, attitudes, values or behaviour. Influence is defined as the *use* of power by means of tactics (e.g. promises, threats, rational persuasion). In general, managers who have power gain compliance from subordinates, respect from superiors and are successful organizational leaders.

The *social power and influence in organizations* research programme integrates organizational psychology and applied social psychology and tackles both basic and applied issues. This programme of research addresses three fundamental questions:

Address for correspondence: Herman Aguinis or Kurt Kraiger, Department of Psychology, University of Colorado at Denver, Campus Box 173, P.O. Box 173364, Denver, Colorado, 80217-3364, USA. Further information regarding the Industrial and Organizational Psychology Programme at the University of Colorado at Denver can be found on the World Wide Web at <http://www.cudenver.edu/public/psych/index.html>.

- What are the factors affecting perceptions of power? (e.g. Aguinis and Adams 1997; Aguinis, Nesler, Quigley and Tedeschi 1994; Nesler, Aguinis, Quigley and Tedeschi 1993)
- What are the factors affecting the choice for various influence tactics? (e.g. Aguinis, Nesler, Hosoda and Tedeschi 1994)
- What are the consequences and correlates of various power bases? (e.g. Aguinis, Nesler, Quigley, Lee and Tedeschi 1996)

Selective recent findings. Recent findings of this programme of research relate to each of the three aforementioned questions. First, regarding power perceptions, a recent study by Aguinis and Adams (1997) investigated two competing paradigms that attempt to explain poor evaluations and slow organizational advancement of female managers as compared to male managers. The *social-role model* attributes slow female organizational advancement to incongruencies between gender-role expectations attributed to women (e.g. indirectness) and behaviours necessary for managerial success (e.g. directness). Alternatively, the *structural model* attributes women's slower upward mobility to work structures and institutional practices such as the distribution of power, rewards and opportunities within the organization. Aguinis and Adams adopted a strong inference epistemological approach and experimentally confronted opposing predictions derived from these models. A sample of 88 non-traditional students (i.e. older, considerable work experience) viewed a videotape of a manager (female, male) occupying *identical* roles in the organizational structure and using one of two forms of influence behaviours (direct, indirect). Thus, Aguinis and Adams assessed the effects of gender role expectations on the evaluations of female managers controlling for the potential confounding effects of organizational role expectations. Overall, the prediction of the *social-role model* was that female managers would receive more negative evaluations than male managers when using (sex role incongruent) direct influence behaviours. On the other hand, the prediction of the *structural model* was that there would be no gender-based differences (because the female and male managers hold identical positions) and there would be a main effect for influence use (because direct influence is more congruent with the managerial position than indirect influence). Supporting the structural model, ratings of (a) managerial power, (b) leadership effectiveness, (c) managerial attributes, and (d) reactions to an influence attempt were affected by the type of influence used and not by the gender of the manager. In sum, this experiment demonstrated that when managerial role and influence use were held constant,

potential gender-based expectations were overridden or neutralized and perceptions of power and other attributes associated with organizational advancement and success were not affected by the gender of the manager.

Second, regarding the choice for influence tactics, Aguinis, Nestler, Hosoda and Tedeschi (1994) conducted an experiment in which 154 undergraduate students were presented with vignettes describing a downward influence situation together with conditions varying (a) the types of jobs of the agent and target of the influence situation, (b) the role identities of the actors and (c) the goal of the actor attempting to influence the target. Confirmatory factor analysis was used to test the hypothesis that the preference order for influence tactics was invariant across situations. As predicted, the likelihood and relative degree to which the influence tactics would be used were similar across situations, thus providing further supportive evidence for the schema-based notion of a 'persuade package.' This persuade package can be defined as a schema, a mental framework regarding the social influence domain that directs the encoding, storage and retrieval of information related to this domain. Thus, this research extended the application of schemata from organizational areas such as performance appraisal and leadership perceptions to the tactics chosen by individuals when faced with an influence situation.

Third, regarding correlates of power perceptions, Aguinis, Nesler, Quigley, Lee, and Tedeschi (1996) utilized the French and Raven (1959) power taxonomy (coercive, expert, legitimate, referent and reward) to investigate graduate students' perceptions of their supervising professor's power, and the relationship between professor's power and various students' perceptions, intentions, and behaviours. Results showed that faculty power bases were related to several variables critical to student satisfaction and success. For instance, *expert power* was positively associated with perceptions of the quality of the relationship, supervisor's trustworthiness, supervisor's credibility, intentions to invite the supervisor to serve and chair a dissertation or thesis committee and intentions to engage in collaborative research with the supervisor. Based on these findings, the power relationship between faculty advisors and graduate students should not be overlooked in future research addressing graduate student education.

Current projects and future directions. CU-Denver faculty, undergraduate, and graduate students are actively involved in continuing to develop the *Social Power and Influence in Organizations* research programme. For example, current projects include the experimental

investigation of the effects of disability status and nonverbal behaviour on power bases perceptions (e.g. Aguinis, Simonsen and Pierce in press; Jordan, Blake and Aguinis 1996); the examination of the effects of gender on perceptions of power (Aguinis, Pierce and Veng 1996) and tests of hypothesized relationships among power, romantic relationships, and sexual harassment in the workplace (e.g. Pierce and Aguinis in press; Pierce, Byrne and Aguinis 1996).

Personnel selection

Primary questions addressed. The traditional area of *Personnel Selection* is a second programme of research at CU-Denver. The overall goal of the research conducted in this area is to improve personnel selection decision making. For instance, this programme of research investigates the accuracy of decisions regarding whether a test predicts performance differentially for various groups (e.g. women and men) In this particular example, differential prediction implies that gender moderates the relationship between test scores and performance and, therefore, the test is biased. The *Personnel Selection* area of research addresses specific questions including:

- What is the effectiveness of moderated multiple regression (MMR) to accurately assess differential prediction and moderating effects? (Aguinis 1995; Aguinis, Pierce and Stone-Romero 1994) What are some of the factors that affect the accuracy of moderated multiple regression? (Aguinis and Stone-Romero in press)
- What are some possible solutions that could be implemented to improve the accuracy of MMR to assess differential prediction and estimate moderating effects? (Aguinis, Bommer and Pierce in press)

Selective recent findings. Moderated multiple regression is a consensually used technique to assess differential prediction (i.e. whether the relationship between pre-employment test scores and job performance scores varies across groups) However, despite its popularity, recent research on the MMR approach to the estimation of differential prediction and moderator variables in general has identified several factors that reduce statistical power below acceptable levels. Consequently, researchers may *erroneously* dismiss theoretical models that include moderated relationships (Aguinis and Stone-Romero in press; Stone-Romero, Alliger and Aguinis 1994) For instance, Aguinis (1995) reviewed factors that affect the statistical power of hypothesis tests conducted using this technique and proposed solutions to low power

situations. Aguinis's review suggests that researchers who use MMR to assess differential prediction, and moderating effects in general, should interpret null moderating effect findings with extreme caution: low statistical power is a frequent, and likely, explanation for the lack of support for moderating effect hypotheses.

Current projects and future directions. Current projects in the *Personnel Selection* research programme include (a) a consideration of the importance of the homogeneity of error variance assumption for differential prediction decisions, (b) the implications of indirect range restriction for validity generalization conclusions and (c) the investigation of banding as an alternative to strict top-down selection.

First, regarding the homogeneity of error (residual) variance, this is a necessary statistical assumption for the appropriate use of MMR for assessing differential prediction (i.e. inequality of regression slopes across moderator-based subgroups). Aguinis and Pierce (1996a) clarified issues pertaining to the violation of the homogeneity of error variance assumption and differentiated it from the homoscedasticity assumption. In addition, they delineated the practical implications of the violation of the homogeneity of error variance assumption in terms of personnel selection decision making. Finally, they critically reviewed solutions recently proposed to mitigate the detrimental effects of the violation of the homogeneity of error variance assumption for the accurate estimation of differential prediction. In general, violating this assumption has consequential implications for personnel decision making: unbiased tests may be *incorrectly* judged as biased and biased tests may be *incorrectly* judged as unbiased.

Second, Aguinis and Whitehead (1996) conducted Monte Carlo simulations to investigate whether indirect range restriction (IRR) on two variables X and Y spuriously increases the sampling error variability in the correlation coefficient between them. The manipulated parameters were (a) IRR on X and Y (i.e. direct restriction on a third variable Z), (b) population correlations ρ_{XY} , ρ_{XZ} , and ρ_{YZ} (ranging from $\rho = 0.10$ to 0.90 in increments of 0.10), and (c) sample size ($N = 60, 100, \text{ and } 140$) Across values of sample size and variable intercorrelations, IRR increased the sampling error variance of r_{XY} to values as high as 10% larger than the analytically-derived expected values. Given these results, Aguinis and Whitehead urged researchers to consider the implications of IRR for validity generalization research: In the presence of IRR, variability across study-level r s can be artificially inflated. Consequently, in IRR situations, researchers need

to make theory-based decisions to ascertain whether the effects of IRR are artifactual (Schmidt, Law, Hunter, Rothstein, Pearlman and McDaniel 1993) or caused by situational-specific moderating effects (James, Demaree, Mulaik and Ladd 1992)

Third, Aguinis, Cortina and Goldberg (1997) advanced and illustrated a new procedure for the formation of equivalence bands in pre-employment tests for situations in which criterion data are available. Cascio, Outtz, Zedeck and Goldstein (1991) proposed the use of test score banding as a way to incorporate both utility and adverse impact considerations in the selection process. Banding is an alternative to the strict top-down selection strategy, which typically leads to adverse impact, and is based on the premise that pre-employment measures are never perfectly reliable. Thus, an observed difference in the scores of two job applicants may be the result of measurement error instead of actual differences in the construct (e.g. skills and abilities) that is measured.

Banding is advocated as a procedure that considers 'all scores falling within the band as relatively equally qualified, within the limits of measurement error' (Cascio *et al.* 1991 p. 242) However, the Cascio *et al.* procedure does not include specific criterion information. Nevertheless, the ultimate goal of banding, and any other decision-making procedure in personnel selection, is to ascertain whether two applicants will *perform* at similar levels on the job. Hiring decisions are then made based on these predictions. However, the Cascio *et al.* model does not explicitly consider the precise predictor-criterion relationship and operates under the assumption that there is an acceptable level of useful empirical or content validity. Accordingly, based on this 'acceptable validity' premise, equivalence regarding predictor scores is equated to equivalence regarding criterion scores.

However, few pre-employment tests explain more than one quarter of the variance in a given criterion. Thus, the assumption that two applicants who are indistinguishable (i.e. falling within the same band) or distinguishable (i.e. not falling within the same band) regarding the predictor construct are also indistinguishable or distinguishable regarding the criterion construct is not tenable.

The banding procedure proposed by Aguinis *et al.* (1997) incorporates the predictor-criterion relationship not specifically addressed in the Cascio *et al.* procedure. This new procedure consists of: (1) the computation of the width of a band of statistically indistinguishable scores on a performance measure *Y* (2) the correction for artifacts of the validity coefficient representing the relationship between the criterion *Y* and the

predictor *X* (3) the determination of the upper and lower limits on the band for *Y* and (4) the computation of the two scores on *X* which produce predicted scores for *Y* equal to the upper and lower limits of the band on *Y*. Thus, this new approach generates two scores on *X* that are associated with a range of indistinguishable scores on the predicted value of the *criterion*.

Applied Research Methods

Primary questions addressed. The *Applied Research Methods* programme includes the investigation of various measurement and applied methodological issues (e.g. Aguinis 1994) Recent and current work in this area investigates the potential, and relative effectiveness, of alternative research methods that may improve the validity of research conclusions in applied psychology. Two such methods are the bogus pipeline (Aguinis, Pierce and Quigley 1993, 1995) and meta-analysis (Aguinis and Pierce 1996b) Also, this area of research includes investigations of similarities and differences between the scientific method of knowledge generation and other models such as action research (Aguinis 1993) and religion (Aguinis and Aguinis 1995).

Selective recent findings. Jones and Sigall (1971) first suggested the bogus pipeline (BPL) as a technique to minimize the social desirability component of traditional self-report measures in social psychological research. The BPL comprises a set of procedures that lead respondents to believe that researchers have a powerful, sophisticated and practically infallible lie detector. In actuality, the supposed lie-detector is not real. Because respondents do not want to be second-guessed by a machine, they may be motivated to provide more honest self-reports.

Applied psychologists have been enthusiastic about the BPL procedure, especially those interested in collecting sensitive information such as attitudes on racial issues. However, despite a large number of studies using the BPL, some findings were supportive of this technique, whereas others indicated that the BPL did not enhance the validity of self-reports. To solve this controversy, Aguinis, Pierce and Quigley (1993) conducted a meta-analysis to test whether the use of self-report measures within the bogus pipeline paradigm yields more valid responses than the use of self-report measures alone for assessing *cigarette smoking* behaviour. The meta-analytic results indicated that, overall, a BPL condition resulted in a larger proportion of subjects reporting that they are frequent smokers, as compared to a self-report measure only (no pipeline) condition. Moreover, tests of categorical models indicated that the enhanced

validity of self-reports within the BPL paradigm is moderated by the following variables: (a) type of BPL presentation employed, (b) type of self-report measure to which the BPL technique is compared and (c) whether most participants are smokers (as indicated by a biochemical marker).

More recently, Aguinis, Pierce and Quigley (1995) conducted two separate meta-analyses to test whether the use of self-report measures within the bogus pipeline paradigm yields more valid responses than self-report measures alone for assessing *alcohol* and *marijuana* consumption. Weighted mean effect sizes (*ds*) were computed to compare the standardized difference between the proportion of self-reported alcohol and marijuana users in BPL groups and no-BPL groups. The obtained *ds* were 0.01 and -0.12 for studies using alcohol and marijuana self-reports, respectively. Nonsignificant chi-square tests, based on *Ns* of 1,892 for the alcohol sample and 1,425 for the marijuana sample, indicated homogeneity of effect sizes for both analyses.

In sum, based on the aforementioned meta-analytic investigations, the BPL methodology seems to be effective at enhancing the validity of self-reported cigarette smoking, but does not seem to be effective at enhancing the validity of self-reported alcohol and marijuana consumption.

Current projects and future directions. Current projects in the applied research methods area include the investigation of ethical issues surrounding the BPL (Aguinis and Handelsman 1997a,b) and the development of new procedures for estimating moderating effects meta-analytically (Aguinis and Pierce, 1996b)

First, regarding the ethical issues surrounding the BPL, it should be noted that despite the existence of more than 60 published studies using the BPL, this literature is surprisingly silent about ethical considerations pertaining to the implementation of this technique. The void in the literature regarding the ethics of the BPL is surprising because the use of the BPL raises ethical issues that go *beyond* those present in more typical deception studies. For example, in more typical deception studies researchers generally mislead subjects by omission or by not revealing the whole truth about the methodology and procedures. On the other hand, researchers who use the BPL do not only conceal the truth, but actively lie to participants not only about the purpose of the study, but about the nature and effectiveness of the 'lie-detector' procedure. Also, participants in BPL studies may feel coerced into revealing sensitive information that is personal, such as marijuana smoking. Because some personal behaviours are illegal, the information gathered in studies using

the BPL may be self-incriminating, and this poses a special threat to participants. The work by Aguinis and Handelsman (1997a,b) provides a conceptual framework that will prove useful in clarifying ethical issues in the use of the BPL.

Finally, regarding the estimation of moderating effects using meta-analysis, research typically compares various approaches and attempts to demonstrate the superiority of a *single* meta-analytic procedure. Moreover, an inspection of meta-analyses published in several I/O psychology journals indicates that researchers also tend to adhere to a *single* meta-analytic strategy. Aguinis and Pierce (1996b) adopted a different perspective and suggested that an *integration* of two major meta-analytic approaches (Hedges and Olkin 1985 and Hunter and Schmidt 1990) is advantageous for detecting moderator variables meta-analytically. Integrating these approaches minimizes Type I statistical error and enhances the validity of conclusions regarding moderating effects. The proposed integrative approach entails the following three steps: (1) statistically correcting for across-study variability in effect size estimates due to methodological and statistical artifacts (2) statistically testing the overall homogeneity of study-level effect size estimates after the artifactual variance has been removed and (3) statistically testing the effects of hypothesized moderator variables. Future simulation work will empirically test the analytic developments and illustrations set forth by Aguinis and Pierce (1996b)

Training Evaluation

Primary questions addressed. Training evaluation is conducted for a number of purposes, including assisting in organizational decision-making (e.g. to retain or eliminate a training programme), providing feedback to trainers and subordinates, or evaluating the effectiveness of new training methods. Training evaluation has become a topic of increasing interest to practitioners as more organizations begin to treat training departments as profit centres and require documentation of the cost effectiveness or return on investment of its training. Despite the theoretical and practical importance of training evaluation, the field of I/O psychology has remained reliant on a circa-1960 evaluation 'model'—Kirkpatrick's (1987) hierarchy. This model posits that training should be evaluated in terms of trainees' reactions to training, trainee learning, changes in on-the-job performance, and changes in organizational effectiveness.

In recent years, several authors have criticized Kirkpatrick's model as atheoretical, incomplete or inaccurate (Alliger and Janak, 1989; Kraiger 1995) From the perspective of facilitating research on

training effectiveness, the harshest criticism is that because the hierarchy is rooted in 1950's behaviourism, Kirkpatrick's model ignores 30 years of research on learning and skill acquisition in its recommendations for how to evaluate learning during training (Kraiger 1995; Kraiger, Ford and Salas 1993)

Consequently, the *Training Evaluation* research programme seeks to integrate current theory and research on human learning and cognition with the practical issues necessitating training evaluation investigations in organizations. This programme of research addresses three fundamental questions:

- Given our current understanding of knowledge and skill acquisition, how should learning during training be evaluated? (e.g. Jonassen, Tessmer and Kraiger 1995; Kraiger *et al.* 1993)
- Given knowledge of instructional objectives and instructional strategies, how can training practitioners best define learning outcomes (upon which evaluation measures should be based) (e.g., Kraiger and Jung 1997)
- How can researchers provide construct-related evidence of the validity of new training evaluation measures (e.g., Jonassen *et al.* 1995; Kraiger and Jung 1997)

Selective recent findings. Kraiger *et al.* (1993) reviewed current research from diverse disciplines such as cognitive psychology, instructional psychology, human factors and human performance to define the ways in which learners acquire knowledge and skills. From this review, they defined three broad classes of learning outcomes: cognitive, affective and skill-based. For each broad class of outcomes, Kraiger *et al.* provided more specific categories of learning outcomes. For example, within the class of cognitive outcomes, Kraiger *et al.* defined declarative, procedural and structural knowledge outcomes. For each outcome, Kraiger *et al.* identified both learning constructs and potentially useful evaluation measures.

More recently, Jonassen *et al.* (1995) extended the Kraiger *et al.* (1993) typology to include nine broad classes and 53 specific learning outcomes. For each specific outcome, Jonassen *et al.* recommended operational measures. Jonassen *et al.* also provided a list of 84 instructional strategies (e.g. providing prototypical examples or corrective/remedial feedback) and provided a conceptual matrix to link their proposed learning outcomes to each instructional strategy. Thus, this matrix provides a mechanism for training researchers to evaluate the content validity or relevance of measures already in use or to plan evaluation measures which are optimally linked to the nature of the instructional intervention. For example, measures of structural knowledge

or semantic models are recommended as the most appropriate forms of instruction that teach categorization or classification within a content domain.

Kraiger and Jung (1997) argue that for typologies such as those offered by Kraiger *et al.* (1993) and Jonassen *et al.* (1995) to be useful to practitioners and training researchers, there must be methods of identifying learning outcomes given knowledge of instructional processes. Specifically, Kraiger and Jung note that the Kraiger *et al.* and Jonassen *et al.* typologies recommend evaluation measures *given knowledge* of learning outcomes, but that learning outcomes may not be clearly defined in many situations. Typically, instructional designers create instructional objectives to guide the training development and criterion development processes, but learning outcomes may remain implicit. In contrast to learning outcomes, which are broad-based expressions of the class of change expected as a result of training (e.g. well-defined goal structures or greater structural knowledge), instructional objectives refer to specific changes in trainees based on the instructional content (e.g. trainees will be able to create a network of goals and subgoals necessary to detect a short in an electric circuit board). Kraiger and Jung provide guidelines for generating learning outcomes given knowledge of the goals of training, the instructional strategies, and the instructional objectives.

Finally, given the number of new measures proposed by Kraiger *et al.* (1993) and Jonassen *et al.* (1995), it is important to establish the construct validity of these new measures. Historically, it has been difficult to establish the construct validity of criterion measures (Kraiger and Jung 1997; Kraiger *et al.* 1993). Consider classic selection research: in contrast to predictor measures, which can be validated against later job performance, there is often no domain-relevant measure of performance to which criterion scores can be compared. However, the opportunity to provide construct-related evidence of the validity of criterion measures is greater in the training evaluation domain than it is in selection research. This is because training researchers can form specific hypotheses about the nature of change on multiple measures given knowledge of the instructional methods, intended learning outcomes and type of trainees. For example, Kraiger, Salas and Cannon-Bowers (1995) investigated the validity of a measure of structural knowledge by not only examining its relationship to a long-term measure of transfer, but also by investigating its sensitivity to a specific instructional technique (use of advanced organizers) that in theory should be related to

changes in structural knowledge. Jonassen *et al.* (1995), Kraiger (1995/1996) and Kraiger and Jung (1997) provide further discussion of research designs that can be used to provide construct-related evidence of the validity of new training evaluation measures.

Current projects and future directions. Current research in the *Training Evaluation* research programme at the CU-Denver include investigating the impact of the Kraiger *et al.* (1993) and Jonassen *et al.* (1995) learning outcomes typology for evaluating training programmes.

As noted above, the validity of any new training evaluation measure can be investigated by examining its sensitivity to training manipulations which should have an effect on that measure (but not necessarily on other measures). Several current studies being conducted by graduate students at the CU-Denver are using this approach to validate new training evaluation measures. Each study uses some form of computer-assisted instruction. In one paradigm, a simple form of an intelligent tutoring system (ITS) for teaching automobile engine diagnostics is being used as a criterion task. Subjects complete either of two forms of pre-training prior to full training on the ITS. One form of training is based on concept learning (Klausmeier 1992): subjects watch a videotape providing an overview of engine systems (with examples of category and noncategory members) and a conceptual overview of how the ITS works. The other form of training is more hands-on or experiential, simulating aspects of situated learning (Glaser and Bassock 1989). In this condition, subjects are given no information of engine systems or the ITS, but learn by trial and error on ITS practice sessions. Learning environments and subject goals and feedback are also manipulated. Four conceptually-based evaluation measures are administered before and after pre-training, and performance on these measures are examined as a function of pre-training conditions (e.g. does a measure of structural knowledge show greater change during concept learning than trial and error?) and as a predictor of actual performance or learning on the ITS.

Another similar paradigm examines pretest/posttest changes on measures of structural knowledge and procedural knowledge on either of two methods of learning word processing. One is a commercially available videotape which is intended to convey conceptual knowledge of menus and functions in a word-processor, while the other is an interactive hyper-text system which is intended to increase subjects' knowledge of word-processing tasks such as formatting, using tools and creating tables. In

this latter series of studies, measures of learning during training (assessed by both types of measures) are compared to both indices of transfer of learning in training and transfer of training to actual performance environments (see Kraiger 1995 for more details)

Finally, the *Training Evaluation* research team is attempting to evaluate the impact of the Kraiger *et al.* (1993) and Jonassen *et al.* (1995) typologies through an extensive meta-analysis of learning outcomes in studies of computer-assisted-instruction (CAI). Although other researchers have conducted meta-analyses of CAI, there have been no prior efforts to match the type of instruction to the type of measure at the level of a meta-analysis. For example, Kulik and Kulik conducted a number of meta-analyses in the late 1980s (e.g. Kulik and Kulik 1986; Kulik, Kulik and Bangert-Drowns 1985), but examined only the type of instruction provided by the computer. In our meta-analysis, we are examining effect sizes based on the type of instruction, e.g. are scores on skills tests greater for forms of computerized instruction that provide practice or are scores on measures of structural knowledge greater when advanced organizers are provided in training?

Cognitive Assessment

Primary questions addressed. An important implication of the learning outcomes typology of Kraiger *et al.* (1993) is that the type of evaluation measure used in a training context should match the various cognitive outcomes hypothesized to occur in training. To be able to conduct this type of evaluation, new methods of assessing cognitive outcomes must be developed and validated. Toward these ends, the *Cognitive Assessment* research programme addresses the following questions:

- How can methods of structural assessment be developed in a way that efficiently *and* validly measures structural knowledge? (e.g. Goldsmith and Kraiger 1997; Kraiger *et al.* 1995)
- How can methods of assessing individuals' knowledge structures be extrapolated to assess team shared knowledge structures and team mental models? (e.g. Kraiger and Wenzel in press)

Selective recent findings. In recent years, cognitive psychologists have begun to focus more extensively on how individuals represent, structure, or organize their knowledge in a particular field of study. The representation of these concepts is referred to as the *knowledge structure* of a domain, and the measurement of that representation is called structural assessment (Goldsmith and Kraiger 1997). One technique

which has been frequently used to assess structural knowledge is called Pathfinder (Schvaneveldt, Durso and Dearholt 1989). Pathfinder transforms pairwise similarity ratings into a network structure in which the relationship among concepts are shown as distance-weighted linkages among nodes. While research supports the validity of Pathfinder representations (e.g. Acton, Johnson and Goldsmith 1994), the technique is time-consuming in its present form.

The success of structural knowledge techniques such as Pathfinder at predicting individual performance has led some researchers to speculate whether these measures would be similarly useful for assessing shared mental models in teams (Jenkins and Rentsch 1995; Kraiger and Wenzel in press). However, because shared mental models in teams are considered to be complex, multi-dimensional constructs (Cannon-Bowers, Salas and Converse 1992, Klimoski and Mohammed 1994), so that more theory and research is necessary before identifying which specific aspects of shared mental models can be assessed via structural assessment, and how those measures should be developed.

Current projects and future directions. Current projects in the *Cognitive Assessment* research programme include: (a) assessing the validity and efficiency of methods of extracting similarity ratings for Pathfinder analyses (b) deriving construct-valid measures of multiple dimensions of shared mental models and (c) measuring shared mental models and cue-strategy relationships in air traffic controllers.

As noted above, submitting pairwise similarity ratings to a Pathfinder analysis has produced valid representations of subjects' structural knowledge. However, the process of collecting pairwise ratings can be time consuming. Generally, at least 20 concepts are necessary to adequately represent a domain (Goldsmith and Kraiger 1997) and collecting pairwise ratings on 20 concepts can take 20–30 minutes. Consequently, one series of studies conducted by the *Cognitive Assessment* research programme has been examining whether the reliability and validity of Pathfinder representations can be maintained while diminishing the time for data collection.

In the first study, we resurrected a technique that had been used effectively in paired comparison studies in the 1950's (e.g. McCormick and Bachus 1952). Subjects provided pairwise similarity ratings for 24 football terms using either a paper-and-pencil format or a computer rating programme. One week later, the subjects rated the same terms again, but half the subjects received a reduced set of terms. For

these subjects, approximately 25% of the pairs were not presented, and their ratings for the missing pairs were predicted from their ratings on other pairs. We found that ratings could be obtained more quickly on the computer than via paper-and-pencil and that rating the reduced set of terms took less time as well. However, the test-retest reliability of the representations were lower for the subjects who received the reduced set of pairwise ratings. Thus, this particular study did not support this strategy for reducing the time for data collection.

In a second series of studies, we are attempting to reduce the rating time by varying the rating scale. Two different task domains are being investigated. In one, subjects rate 20 word processing terms (e.g. File, Save As) using either a two-point, four-point or nine-point similarity scale. In another experiment, subjects rate poker terms (e.g. ante, flush). After individual ratings are analyzed using Pathfinder, we score each subject by comparing the similarity of their representations to an expert solution. We will then examine the validity of their representations by correlating the similarity scores with scores on either a ten-item formatting and editing task (e.g. placing text in bold, move text, save files) or a 15-item decision-making task (e.g. deciding whether to call or fold given a certain hand). We are hypothesizing that the two- and four-point scales will take less time than the nine-point scales (because fewer distinctions in similarity were necessary). Also, we are investigating whether there will be a decrement in validity for the two- or four-point scales and we are determining whether these effects are consistent across well-structured (word-processing) or ill-structured (poker) domains.

A second programme of research within the *Cognitive Assessment* area is attempting to develop and validate measures of shared mental models in teams. Early attempts to measure shared mental models focused on single measures of unidimensional constructs (e.g. Jenkins and Rentsch 1995; Minionis, Zaccaro and Perez 1995). However, Kraiger and Wenzel (in press) argued that since the shared mental model construct is assumed to be multi-dimensional (Cannon-Bowers *et al.* 1993), its measures should be multi-dimensional as well. To demonstrate the efficacy of a multi-dimensional approach, we recently had both intact and non-intact teams complete three measures of aspects of shared mental models (assessing declarative knowledge, procedural knowledge and structural knowledge), and then complete three separate tasks (Kraiger, Krause and Hart 1996). Each task emphasizes a unique type of knowledge. The tasks and knowledge measures concern an artificial task such as

planning an outdoor dinner party. For example, the structural knowledge measure examines the similarity of Pathfinder representations of domain-specific terms (e.g. cooking, marinating), and the related task requires each partner to describe what they would do if they wanted to support their partner's preparation in the areas of cleaning, cooking or planning. Consistent with our multi-dimensional approach, we found that team scores on the knowledge tests were more highly correlated with scores on the matched tasks than on unmatched task. A follow-up study is planned to collect similar data from a real world setting.

A related line of research is attempting to develop and validate measures of shared mental models (for teams and tasks), shared expectations, and cue-strategy associations in teams of air traffic controllers. One challenge in this research is that the measures of these constructs must be relatively brief as they are being included in an assessment battery that includes a number of other individual- and team-level measures of knowledges (e.g. task knowledge, inter-positional knowledge) and attitudes (e.g. collective efficacy, mutual trust). The measures are being validated in two contexts. One is an experimental context in which expert and novice controllers complete the measures and then perform a complex PC-based air traffic control simulation. The research question focuses on the extent to which the measures individually and collectively predict team performance on the simulation. The second context is a series of field studies in which the measures are treated as dependent variables and characteristics of team demography (e.g. years experience, previous experience as pilot) and the organization (e.g. level of traffic, facility leadership style) are used as predictors.

Center for Applied Psychology

Many of the studies that comprise the aforementioned five research programmes were completed under the auspices of the *Center for Applied Psychology* (CAP). The mission of CAP is to assist client organizations improve overall productivity and individual effectiveness through the application of state-of-the-art knowledge and research. CAP provides a means of linking faculty, graduate students and resources of the University of Colorado at Denver to local, out-of-state and overseas businesses and agencies. These links may come in the form of consulting, collaborative research, seminars and training or other applications determined by client needs.

In line with the scientist-practitioner model advocated by the Society for Industrial and

Organizational Psychology (SIOP), CAP provides CU-Denver faculty and students with institutional and material support to conduct applied organizational research. CAP, directed by Professor Kurt Kraiger, sponsors projects in all areas of I/O psychology. In the last two years, CAP has entered partnerships with a number of organizations providing consultative services, sponsoring programming or conducting applied research.

Four illustrations of ongoing activities sponsored by CAP include: (1) service audits for an international consulting firm and a veterinary hospital (2) programme evaluation for the Youth Services Program, State of Colorado (3) hosting a 3-day conference on distance learning and multimedia education and (4) evaluating the effectiveness of a restructuring of the civil engineering curriculum at the United States Air Force Academy. In addition, CAP is actively engaged in developing contacts in Latin America in order to implement I/O psychology interventions in emerging countries (Aguinis and Kraiger 1996).

Several highlights: the service audits are coordinated efforts to review an organization's capacity to provide service and provide recommendations for improving service. Activities include observations and interviews with staff or customers, literature reviews within the industry of the organization and other strategies as necessary. (For example, for the service audit of the veterinary hospital, CAP staff conduct a survey on the Internet of clients' service preferences when choosing a new veterinarian), the evaluation of the Air Force Academy's civil engineering programme has been a three-year effort to examine the impact of a 'construct first, design later' method of instruction on student attitudes and performance in the classroom, as well as on performance as an officer in one's first assignments.

Summary

The Industrial and Organizational Psychology Programme at CU-Denver has the primary objective of training students to perform psychological research, evaluation and services in public or private sector organizations. CU-Denver I/O faculty, undergraduate and graduate students are actively engaged in several programmes of research. This article briefly summarized the characteristics and goals of the I/O programme, as well as five ongoing, programmatic research agendas: (a) Social Power and Influence in Organizations, (b) Personnel Selection, (c) Applied Research Methods, (d) Training Evaluation, and (e) Cognitive Assessment. Finally, the Center for Applied

Psychology provides a propitious institutional environment for the further development of what we believe are engaging and provocative research agendas.

References

- Acton, W.H., Johnson, P.J. and Goldsmith, T.E. (1994) Structural knowledge assessment: comparison of referent structures. *Journal of Educational Psychology*, **86**, 303–311.
- Aguinis, H. (1993) Action research and scientific method: presumed discrepancies and actual similarities. *Journal of Applied Behavioral Science*, **29**, 416–431.
- Aguinis, H. (1994) A QuickBASIC program for generating correlated multivariate random normal scores. *Educational and Psychological Measurement*, **54**, 687–689.
- Aguinis, H. (1995) Statistical power problems with moderated multiple regression in management research. *Journal of Management*, **21**, 1141–1158.
- Aguinis, H. and Adams, S.R. (1997) *Social-role versus structural models of gender and influence in organizations: a strong inference approach*. Manuscript submitted for publication.
- Aguinis, H. and Aguinis, M. (1995) Integrating psychological science and religion. *American Psychologist*, **50**, 541–542.
- Aguinis, H., Bommer, W.H. and Pierce, C.A. (1996) Improving the estimation of moderating effects by using computer-administered questionnaires. *Educational and Psychological Measurement* **6**, 1045–1049.
- Aguinis, H., Cortina, J.M. and Goldberg, E. (1996) A new procedure for computing equivalence bands in personnel selection. Manuscript submitted for publication.
- Aguinis, H. and Handelsman, M.M. (1997a) Ethical issues in the use of the bogus pipeline. *Journal of Applied Social Psychology*, **27**, 557–573.
- Aguinis, H. and Handelsman, M.M. (1997b) The unique ethical challenges of the bogus pipeline methodology: let the data speak. *Journal of Applied Social Psychology*, **27**, 582–587.
- Aguinis, H. and Kraiger, K. (in press) Practising what we preach: competency-based assessment of industrial/organizational psychology graduate students. *The Industrial-Organizational Psychologist*.
- Aguinis, H. and Kraiger, K. (1996) Intervening in emerging markets: back to the basics, or go home empty-handed. *The Industrial-Organizational Psychologist*, **33**(3), 84–87.
- Aguinis, H., Nesler, M.S., Hosoda, M. and Tedeschi, J.T. (1994) The use of influence tactics in persuasion. *Journal of Social Psychology*, **134**, 429–438.
- Aguinis, H., Nesler, M.S., Quigley, B.M. and Tedeschi, J.T. (1994) Perceptions of power: a cognitive perspective. *Social Behavior and Personality*, **22**, 377–384.
- Aguinis, H., Nesler, M.S., Quigley, B.M., Lee, S. and Tedeschi, J.T. (1996) Power bases of faculty supervisors and educational outcomes for graduate students. *Journal of Higher Education*, **67**, 267–297.
- Aguinis, H. and Pierce, C.A. (1996a, April) Heterogeneity of error variance and differential prediction: clarifications, implications, and solutions. In M.J. Burke (Chair), *Quantitative Advances in Personnel Selection Research and Practice*. Symposium conducted at the meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Aguinis, H. and Pierce, C.A. (1996b) Testing moderator variable hypotheses meta-analytically. Manuscript submitted for publication.
- Aguinis, H., Pierce, C.A. and Quigley, B.M. (1993) Conditions under which a bogus pipeline procedure enhances the validity of self-reported cigarette smoking: A meta-analytic review. *Journal of Applied Social Psychology*, **23**, 352–373.
- Aguinis, H., Pierce, C.A. and Quigley, B.M. (1995) Enhancing the validity of self-reported alcohol and marijuana consumption using a bogus pipeline procedure: A meta-analytic review. *Basic and Applied Social Psychology*, **16**, 515–527.
- Aguinis, H., Pierce, C.A. and Stone-Romero, E.F. (1994) Estimating the power to detect dichotomous moderators with moderated multiple regression. *Educational and Psychological Measurement*, **54**, 690–692.
- Aguinis, H., Pierce, C.A. and Veng, L.M. (1996 April) Consequences of female managerial influence style. Paper presented at the meeting of the Rocky Mountain Psychological Association, Park City, UT.
- Aguinis, H., Simonsen, M.M. and Pierce, C.A. (in press) *Effects of nonverbal behavior on perceptions of power bases*.
- Aguinis, H. and Stone-Romero, E.F. (in press) Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology*.
- Aguinis, H. and Whitehead, R. (1996) *Sampling variance in the correlation coefficient under indirect range restriction: implications for validity generalization*. Manuscript submitted for publication.
- Alliger, G.M. and Janak, E.A. (1989) Kirkpatrick's levels of training criteria: thirty years later. *Personnel Psychology*, **42**, 331–342.
- Cannon-Bowers, J.A., Salas, E. and Converse, S. (1993) Shared mental models in expert team decision making. In N.J. Castellan, Jr. (ed.), *Current Issues in Individual and Group Decision Making* (pp. 221–246) Hillsdale, N.J.: Erlbaum.
- Cascio, W.F., Outtz, J., Zedeck, S. and Goldstein, I.L. (1991) Statistical implications of six methods of test score use in personnel selection. *Human Performance*, **4**, 233–264.
- Glaser, R. and Bassock, M. (1989) Learning theory and the study of instruction. *Annual Review of Psychology*, **40**, 631–666.
- Goldsmith, T. and Kraiger, K. (1997) Applications of structural knowledge assessment to training evaluation. In J.K. Ford, S.W.J. Kozlowski, K. Kraiger, E. Salas, and M.S. Teachout, (eds.) *Improving Training Effectiveness in Work Organizations* (pp. 73–96) Mahwah, NJ: Lawrence Erlbaum.
- Hedges, L.V. and Olkin, I. (1985) *Statistical Methods for Meta-analysis*. Orlando, FL: Academic Press.

- Hunter, J.E. and Schmidt, F.L. (1990) *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage.
- James, L.R., Demaree, R.G., Mulaik, S.A. and Ladd, R.T. (1992) Validity generalization in the context of situational models. *Journal of Applied Psychology*, **77**, 3–14.
- Jenkins, N.M. and Rentsch, R.R. (1995, May) *The effects of teamwork schema similarity on team effectiveness and fairness perceptions*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Jonassen, D., Tessmer, M. and Kraiger, K. (1995) A plan for research, design, and evaluation of intelligent tutoring systems. Paper submitted as Air Force Technical Paper.
- Jones, E.E. and Sigall, H. (1971) The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, **76**, 349–364.
- Jordan, G.A., Blake, A.D. and Aguinis, H. (1996, April) *Implications of the Americans with Disabilities Act for personnel selection*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Kirkpatrick, D.L. (1987) Evaluation of training. In R.L. Craig (ed.), *Training and Development Handbook: a Guide to Human Resource Development*. (3rd ed., pp. 301–319) New York: McGraw-Hill.
- Klausmeier, H.J. (1992) Concept learning and concept teaching. *Educational Psychology*, **27**, 267–286.
- Klimoski, R.J. and Mohammed, S. (1994) Team mental models: construct or metaphor? *Journal of Management*, **20**, 403–427.
- Kraiger, K. (1995, August) *Paradigms lost: applications and misapplications of cognitive science to the study of training*. Invited address at the annual meeting of the American Psychological Association, New York, NY.
- Kraiger, K. (1995/1996) Integrating training research. *Training Research Journal*, **1**, 5–17.
- Kraiger, K., Ford, J.K. and Salas, E. (1993) Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation [monograph]. *Journal of Applied Psychology*, **78**, 311–328.
- Kraiger, K. and Jung, K.M. (1997) Linking training objectives to evaluation criteria. In M. Quiñones and A. Ehrenstein (eds.), *Training for 21st Century Technology: Applications of Psychological Research*. (pp.151–179) Washington, D.C.: APA.
- Kraiger, K., Krause, J.R. and Hart, P.F. (1996, August) *Construct validation of multiple measures of shared mental models*. Paper presented at the annual meeting of the American Psychological Association, Toronto, Canada.
- Kraiger, K., Salas, E. and Cannon-Bowers, J.A. (1995) Cognitively-based measures of learning during training. *Human Factors*, **37**, 804–816.
- Kraiger, K., and Wenzel, L.H. (in press) A framework for understanding and measuring shared mental models of team performance and team effectiveness. In M.T. Brannick, E. Salas, and C. Prince (eds.), *Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Hillsdale, NJ: Erlbaum.
- Kulik, C.-L.C. and Kulik, J.A. (1986) Effectiveness of computer-based education in colleges. *Association of Educational Data Systems Journal*, **19**, 81–108.
- Kulik, J.A., Kulik, C.-L.C. and Bangert-Drowns, R.L. (1985) Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior*, **1**, 59–74.
- McCormick, E.J. and Bachus, J.A. (1952) Paired comparison ratings: I. The effect on ratings of reductions in the number of pairs. *Journal of Applied Psychology*, **36**, 123–127.
- Minionis, D.P., Zaccaro, S.J. and Perez, R. (1995, May) *Shared mental models, team coordination, and team performance*. Paper presented at the meeting of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Nesler, M.S., Aguinis, H., Quigley, B.M. and Tedeschi, J.T. (1993) The effect of credibility on perceived power. *Journal of Applied Social Psychology*, **23**, 1407–1425.
- Pierce, C.A. and Aguinis, H. (in press) Bridging the gap between romantic relationships and sexual harassment in organizations. *Journal of Organizational Behavior*.
- Pierce, C.A., Byrne, D. and Aguinis, H. (1996) Attraction in organizations: a model of workplace romance. *Journal of Organizational Behavior*, **17**, 5–32.
- Ragins, B.R. and Sundstrom, E. (1989) Gender and power in organizations: a longitudinal perspective. *Psychological Bulletin*, **105**, 51–88.
- Schmidt, F.L., Law, K., Hunter, J.E., Rothstein, H.R., Pearlman, K. and McDaniel, M. (1993) Refinements in validity generalization methods: implications for the situational specificity hypothesis. *Journal of Applied Psychology*, **78**, 3–12.
- Schvaneveldt, R.W., Durso, F.T. and Dearholt, D.W. (1989) Network structures in proximity data. In G.H. Bower (ed.), *The Psychology of Learning and Motivation* (Vol. **24**, pp. 249–284) New York: Academic.
- Stone-Romero, E.F., Alliger, G.M. and Aguinis, H. (1994) Type II error problems in the use of moderated multiple regression for the detection of moderating effects of dichotomous variables. *Journal of Management*, **20**, 167–178.