Convergence of Self-Report and Archival Crash Involvement Data: A Two-Year Longitudinal Follow-Up

Winfred Arthur, Jr., Suzanne T. Bell, and **Bryan D. Edwards,** Texas A&M University, College Station, Texas, **Eric Anthony Day,** University of Oklahoma, Norman, Oklahoma, and **Travis C. Tubre** and **Amber H. Tubre,** Texas A&M University, College Station, Texas

This study constructively extends Arthur et al. (2001) by assessing the convergence of self-report and archival motor vehicle crash involvement and moving violations data in a 2-year longitudinal follow-up. The relationships among these criteria, conscientiousness, and driving speed were also assessed using both predictive and postdictive criterion-related validation designs. Data were collected from a 2-year follow-up sample of 334 participants. Results suggested a lack of convergence between self-report and archival data at both Time 1 and Time 2. In addition, the predictor/criterion relationships varied across research design and data source. An actual application of our findings is that the interpretation of relationships between specified predictors and crash involvement and moving violations must be made within the context of the criterion-related validation design and criterion data source. Specifically, predictive designs may produce results different from those of postdictive designs (which are more commonly used). Furthermore, self-report data appear to include a broader range of incidents (more crashes and tickets), and thus researchers should consider using self-report data when they are interested in including lower threshold crashes and tickets that may not be reported on state records (e.g., because of the completion of a defensive driving course).

INTRODUCTION

Using a relatively large and representative sample, the present study constructively replicates and extends Arthur et al. (2001) by investigating the convergence between self-report and archival data and, also, the relationships between these criteria and specified predictors using both predictive and postdictive criterion-related validation designs. The value and importance of constructive replication (Hendrick, 1990; Kelly, Chase, & Tucker, 1979), defined as "research that tests the same hypothesized relationships among the same theoretical constructs as a given earlier study but varies the 'operationalization' of those constructs" (Eden, 2002, p. 842), resides in its ability to strengthen confidence in and enhance the credibility of generalizations when the hypothesized relationships are confirmed using dif-

ferent methods. Along these lines, one of the critical goals of this study was to compare the extent to which predictive and postdictive designs yield similar (or different) criterion-related validities.

There is a reasonably large body of research examining the predictive validity of individual differences in motor vehicle crash involvement (for reviews see Arthur, Barrett, & Alexander, 1991; Elander, West, & French, 1993; Guastello, 1993), and in this research self-report accounts of crash involvement have been the most commonly used criterion (Ball & Owsley, 1991). However, although several reasons may explain the prevalence of self-reports (Elander et al., 1993), questions about their veracity and accuracy have been raised (see Elander et al., 1993; Harano, Peck, & McBride, 1975; Loftus, 1993). Consequently, archival data have been considered as an alternative to self-reports.

Self-Report and Archival Data

For several reasons, self-report measures of crash involvement are the most commonly used criterion measures. For instance, they are relatively easy to collect and enable researchers to sample all crash types. However, self-report measures have been criticized for being inherently susceptible to cognitive and communicative biases, such as selective memory and question comprehension (Elander et al., 1993; Harano et al., 1975; Loftus, 1993; Schwarz, 1999). In contrast, archival records of crash involvement – which are usually obtained from state law enforcement agencies and insurance companies – are assumed to be free of the biases associated with self-report measures. However, this type of data carries its own set of disadvantages. For example, it is generally more difficult to obtain. Another major disadvantage of archival driving records is that they are often incomplete because (a) police selectively report crash involvement, (b) moving violation tickets are removed from drivers' records in exchange for driving course credit, (c) the parties involved sometimes choose not to report the crash, or (d) interstate information exchange of data on crashes and moving violations is incomplete (Burns & Wilde, 1995; McGuire, 1973; Smith, 1976).

Given the important trade-offs between using either archival or self-report data in motor vehicle crash research, few studies have investigated the convergence between these two sources of criterion data. A handful of studies have collected crash data via both self-reports and state records (e.g., Dalziel & Job, 1997; Hilakivi et al., 1989; Marottoli, Cooney, & Tinetti, 1997; Owsley, Ball, Sloane, Roenker, & Bruni, 1991), and some have included performance on a driving simulator (e.g., Szlyk, Alexander, Severing, & Fishman, 1992; Szlyk, Fishman, Severing, Alexander, & Viana, 1993; Szlyk, Seiple, & Viana, 1995). However, most of these studies have used very restricted and specialized samples (e.g., older adults or participants with specialized medical conditions, such as retinitis pigmentosa and juvenile macular dystrophies). In addition, the convergence between the criterion sources has typically not been assessed because it was not the primary focus of the study. McGuire (1973), McGwin, Owsley, and Ball (1998), and Smith (1976) are three examples of studies that specifically investigated the convergence issue.

In addition to assessing the convergence between archival and self-report data, McGuire (1973) investigated the presence of systematic bias in official driving records. Archival driving data were obtained from the Mississippi Highway Patrol, and self-report data were obtained by telephone interviews. McGuire's (1973) results indicated that crash and citation data were grossly underreported in the official records. Even more important, however, there were systematic sex, race, age, and occupation biases.

McGwin et al. (1998) investigated crash information source as a potential explanation for the inconsistencies observed in the extant literature on medical and functional risk factors for older drivers. McGwin et al. (1998) concluded that there was a moderate level of agreement between self-report and state-recorded crash involvement. However, there were significant differences between crash-involved drivers identified via state records and self-report with respect to factors such as age, race, amount of driving, and vision impairment.

Finally, in a test of Zylman's (1972) claim that research studies based on archival data may yield spurious results, Smith (1976) investigated whether or not the exclusive use of official driver records is appropriate. In addition to reviewing official records obtained from the Road Traffic Authority (Western Australia), he interviewed participants to obtain self-report data. His results indicated that participants self-reported more crashes and driving-related convictions than were present in the official records. In addition, relationships between crashes and specified driver characteristics were significant only when archival data were combined with self-reports, leading him to conclude that his "study provides empirical support for Zylman's (1972) assertion that when used for research purposes, official driver records may yield spurious results" (Smith, 1976, p. 210).

In summary, the preceding studies suggest generally low levels of convergence between self-report and archival data, with the frequency of the former being higher than the latter. Nevertheless, although they provide some preliminary insight about the convergence (or lack thereof) between archival and self-report crash data,

certain limitations of these studies warrant additional research. Smith's (1976) sample was limited to 20- to 23-year-old men; McGwin et al.'s (1998) sample was limited to drivers over the age of 50; and McGuire (1973) did not investigate the stability of predictor/criterion relationships across criterion types. In addition, the limited amount of convergence research has not precluded the general assumption that archival and other "objective" criterion data are superior to self-reports of crash involvement.

In contrast to this general impression, Arthur et al. (2001) concluded that self-report data are not inherently inferior to archival data. Using a postdictive design, they found low agreement between the two criterion sources, with participants self-reporting substantially larger numbers of crashes and moving violations than were reported on their state (Texas) records. Furthermore, different and stronger predictor/criterion relationships were obtained for self-report than for archival data. To assess the stability of Arthur et al.'s (2001) findings, we implemented a 2-year longitudinal predictive follow-up. Thus the first objective of the present study was to replicate Arthur et al.'s (2001) findings concerning the lack of convergence between self-report and archival data sources of motor vehicle crash involvement and moving violations using a 2-year longitudinal follow-up data set.

Predictive and Postdictive Criterion-Related Validation Designs

Criterion-related validity involves an empirical demonstration of predictor/criterion relationships. Predictive, concurrent, and postdictive designs are three types of criterion-related validation designs that differ in terms of the time frames in the collection of the predictor and criterion scores. (Concurrent designs are characterized by the collection of the predictor and criterion scores at about the same point in time. By their very nature, concurrent designs do not lend themselves to use in crash involvement research and therefore are not used in this domain.) Postdictive designs, in which the predictor data are collected after the occurrence of the criteria, are the most commonly used designs in the crash involvement literature (see the meta-analysis by Arthur et al., 1991). Thus, for example, in these designs participants complete the predic-

tor measures, and then the number of crashes in which they have been involved over a specified time interval (e.g., the past 5 years) are collected from either archival data or self-report.

The popularity of postdictive designs can be partially attributed to the nature of crashes, traffic violations, and other similar criteria. Specifically, these criteria are low-frequency base rate incidents that require relatively large samples to yield meaningful and useful data. They also require relatively long periods of time for their manifestation. So, an obvious advantage of postdictive designs is that they allow for a more expeditious collection of data – they are more practical and feasible. They also lend themselves to longer criterion time frames (e.g., number of crashes in the past 5 years).

However, the principal disadvantage of postdictive designs is that they assume that the predictor variables are temporally stable and are not subsequently influenced by involvement in crashes or receiving tickets. For example, in investigating the relationship between risk taking and crash involvement, a postdictive relationship assumes that one's degree of risk taking at the time the crash occurred is the same as when the predictor data are collected. However, it is conceivable that involvement in crashes as a result of risk taking may cause one to subsequently become less of a risk taker. In this case, risk taking is temporally unstable and one's level of risk taking at the time of the crash would not be the same as when one's level of risk taking is assessed. Finally, postdictive designs often suffer from substantial restriction of range on both the predictor and criterion because extremely poor performers are eliminated through mortality and/or attrition (Arthur, Barrett, & Doverspike, 1990). Thus the individuals of most interest are not available in the data.

In contrast, with predictive designs, predictor data are collected before the criterion data. Thus predictor data are collected first and then, after some length of time (e.g., 5 years), crash data are collected. Although the use of predictive designs addresses the methodological problems that characterize postdictive designs, these research designs are comparatively more resource demanding and labor intensive, and they require several years to implement. They require an appreciable time interval between the collection

of predictor scores at Time 1 and criterion data at Time 2 (Arthur et al., 1990). Furthermore, predictive designs (especially with longer time intervals) can suffer from attrition effects on the criterion because of the difficulty in locating participants at the time of data collection. Attrition effects result in a loss of statistical power and introduce internal validity threats. Issues relating to statistical power can be offset by using large samples. Therefore, well-designed predictive studies typically require larger samples than do studies using postdictive designs. In spite of these disadvantages, predictive designs may be more interpretable, especially with attitudinal and dispositional predictors that may be influenced by the occurrence of the criteria (e.g., locus of control and making postevent attributions after being involved in a crash or obtaining a ticket; Arthur & Doverspike, 1992).

As previously noted, Arthur et al. (2001) obtained different predictor/criterion relationships for self-report and archival data, with stronger effects obtained for self-report data. However, these data were postdictive: The criteria represented incidents that had occurred in the 6 years preceding the collection of the predictor scores. Consequently, given the potential methodological problems associated with postdictive designs, the second objective of the present study was to investigate the stability of Arthur et al.'s (2001) predictor/criterion relationships using a 2-year longitudinal predictive design. Accordingly, we compared the extent to which predictive and postdictive designs yielded similar (or different) criterion-related validities.

Predictors

The predictors used in the present study were the same as those used in Arthur et al. (2001) – namely, driving speed and conscientiousness. Conscientiousness has been shown to be an important predictor of motor vehicle crashes (Arthur & Doverspike, 2001; Arthur & Graziano, 1996). The theoretical basis for this relationship is based on the premise that highly conscientious individuals are responsive to social obligation norms (Hogan, 1983) – that is, they are people who report that they comply with the constraint of laws and public regulations, engage in less risk taking, and avoid dangerous situations (however, see Tomlinson-Keasey & Little,

1990). Conscientiousness has special theoretical status in many different accounts of social behavior and personality structure. Relative to the other four dimensions of the five-factor model (FFM), conscientiousness is probably the one that is most associated with differences in adherence to performance standards (Digman & Inouye, 1986; Graziano & Ward, 1992). This conceptual analysis is consistent with Elander et al.'s (1993) review, in which the authors described "thoroughness" (a facet of conscientiousness) as a dimension that "has been consistently related to crash rates independent of age, sex, and mileage" (p. 288).

Unlike conscientiousness, the other four FFM dimensions are less clearly conceptually related to driving crash involvement. However, if Hansen's (1989) hypotheses about links between general social maladjustment (e.g., hostility, anger, self-centeredness) and crash involvement are valid, then there may be links between agreeableness and crash involvement. In addition, Arthur and Graziano (1996) obtained limited support for the relationship between agreeableness and moving violation tickets. Beyond this, we expected no specific relations among the other four dimensions and crashes. However, given the very limited research on the FFM and crash involvement, we investigated possible relationships for exploratory purposes. In addition, what limited FFM/crash research there has been (e.g., Arthur & Doverspike, 2001; Arthur & Graziano, 1996; Arthur et al., 2001) has all been postdictive. Consequently, an additional contribution of the present study to the extant literature is to investigate the relationship between the FFM dimensions and crash involvement using a 2-year predictive design.

The second predictor used in present study was driving speed. Driving speed has been shown to be one of the most important predictors of crashes (Wasielewski, 1984; West, French, Kemp, & Elander, 1993; Wilson & Greensmith, 1983). This relationship has been obtained with different operationalizations of driving speed, including observations, video simulations, and self-reports (e.g., Horswill & McKenna, 1999; West et al., 1993). In addition, West et al. (1993) reported high agreement between self-reports and observational measures of driving speed. The importance of driving speed is reflected in

the fact that exceeding the speed limit or driving at unsafe speeds was the most common driver "error" in fatal crashes in 2001 (National Safety Council, 2002), and driving fatalities have been reported to have increased by almost 30% in the late 1980s in states that raised their speed limits from 55 to 65 miles/hr (from 88.5 to 104.6 km/hr; Sleek, 1996).

METHOD

Participants

We recruited 403 participants from Texas A&M University and its community by advertisements in the school newspaper, local newspapers, and TV stations as well as posted notices both on campus and in the community (e.g., libraries, grocery stores, and driver's license renewal offices). Individuals were required to have a valid Texas driver's license. Complete data were available for 394 participants; this constituted the sample size for Time 1 data. Of the 394 participants from Time 1, 393 had data available from the Texas Department of Public Safety (DPS) at Time 2 (i.e., 2 years after Time 1), and we were able to locate 347 (i.e., 88%) for the 2-year follow-up phone interview. Of these, 2 participants were deceased and 11 refused to participate in the follow-up data collection.

The analyses and results were limited to participants we were able to contact and who were willing to respond in the follow-up interview. Thus the final sample size used in the study was 334. The mean age of these participants at Time 2 was 37.91 years (SD = 17.14, range = 17–87), and the mean number of years they had been driving at Time 2 was 21.22 years (SD = 16.14, range = 3–72). During the Time 1 data collection, 34% of the 334 participants were recruited from the Psychology Department's subject pool and received course credit for participating in the study; the other participants were paid either $20 (28%) or $50 (38%). (The participant fee was increased to $50 when it turned out to be quite difficult to recruit participants at the $20 rate.) No additional compensation was given at Time 2.

Measures

Data were collected from participants on (a) the FFM personality dimensions, (b) typical driv-

ing speed above or below posted speed limits, and (c) the number of defensive driving courses taken. Archival data and self-reports of crash involvement and moving violations served as the criteria. The archival driving record also contained information on the number of defensive driving courses taken.

Goldberg's 100 Unipolar Markers (Goldberg, 1992). This is a standardized measure of the FFM personality dimensions. Using a 9-point scale (1 = *extremely inaccurate,* 9 = *extremely accurate*), participants rated a list of 100 common human traits in terms of how accurately the traits described them. Each factor was measured by 20 items, and for purposes of the present research, a participant's score on each of the five factors was the sum of his or her ratings on the 20 items that represented that factor after the appropriate items had been reverse scored. Internal consistency estimates obtained for the FFM dimension scores were .88 for conscientiousness, .87 for agreeableness, .87 for emotional stability, .89 for extraversion, and .89 for openness to experience.

Driving behavior questionnaire (Arthur & Doverspike, 1992). This measure asked participants to "list the total number of crashes in each of the years listed" on the questionnaire. The listed years were 1987 through 1998. To maintain consistency with the archival data (described later), we limited the analyses to incidents that occurred in the period of 1997 through 1998. These were then summed to create the "self-report crashes" variable. Self-report moving violations data were also reported by year. Test-retest reliabilities of .96 to .98 have been reported for crash and ticket data collected using this questionnaire (Arthur & Doverspike, 1992; Arthur & Graziano, 1996). The driving behavior questionnaire was also used to obtain demographic information such as age, sex, and number of years driving. Self-report data for the typical driving speed above or below posted speed limits and the number of defensive driving courses taken were also obtained via this measure.

A second, shorter version of the driving behavior questionnaire was created for the follow-up and was administered via a telephone interview. This version collected demographic information along with information on the total number of

vehicular crashes, moving violation tickets, and defensive driving courses participants had taken in 1999 through 2000 (i.e., during the 2 years after the original data collection).

Archival driving records. With participants' consent, two sets of archival driving records (Time 1 and Time 2) were obtained from the Texas DPS. These individual driving records contained information on the total number of motor vehicle crashes and moving violation tickets. The 2-year archival record (Time 1) also listed the number of defensive driving courses the participant had taken in the 2-year span. The DPS data at Time 1 spanned 1997 through 1998 (2-year archival record), and the DPS Time 2 data were limited to the years 1999 through 2000 to be consistent with the Time 2 self-reports.

Procedure

At Time 1, participants completed the Goldberg 100 Unipolar Markers and the driving behavior questionnaire. A testing session lasted about 2 hr, and all participants completed the measures in the order presented, as previous research using the same measures has shown an absence of order effects (Arthur & Graziano, 1996). Participants also provided contact information for an alternative source, such as a family member or friend who could reach the participant in the event that the participant's address or telephone number that we had was no longer valid at the 2-year follow-up. Driving records were requested from the Texas DPS.

Two years after the first set of data were collected, participant driving records were obtained from the Texas DPS. In addition, attempts were made to contact all 394 participants from Time 1 for a follow-up interview; 347 were successfully contacted. Once they were contacted, the researchers identified themselves with the study, and those who agreed to participate completed the short form of the driving behavior questionnaire.

RESULTS

The descriptive statistics for the study variables, correlations, and standardized mean differences between the self-report and archival data presented in Table 1 indicate that the convergence between self-report and archival data

was weak for both Time 1 and Time 2 data. Although the correlations between the two criteria were moderately large, participants appeared to self-report substantially more crashes – Time 1, $t(333) = 5.50$, $p < .001$, $d = .38$; Time 2, $t(333) = 7.24$, $p < .001$, $d = .54$ – and tickets – Time 1, $t(333) = 6.97$, $p < .001$, $d = .49$; Time 2, $t(333) = 8.52$, $p < .001$, $d = .67$ – than were present in the Texas DPS records. Concerning the magnitudes of these effects, it is worth noting that Cohen (1992) described ds of .20, .50, and .80 as small, medium, and large, respectively. Similar differences were found for the number of self-report and DPS defensive driving courses assessed at Time 1, $t(333) = 10.67$, $p < .001$, $d = .82$. Overall, for the Time 1/Time 2 differences, the degree of agreement between the two sources was lower for the Time 2 follow-up data than it was for the Time 1 data (although the overall difference was not statistically significant).

We assessed the effect of data source by comparing self-report data and archival data. Participants were cross-classified with respect to self-reported and state-reported incidents for crashes and moving violations, thus allowing the creation of four mutually exclusive groups: (a) no self-report or state report, (b) self-report but no state report, (c) no self-report but state report, and (d) self-report and state report (Table 2). These data indicate that for all criteria and for both the Time 1 and Time 2 data, the lack of agreement was attributable to the number of participants (e.g., 12%–24%) in the self-reported but not state-reported category.

Table 3 presents the predictor/criterion correlations, and the results reflect modest differences in the predictor relationships for the two criterion types. The postdictive validities for conscientiousness were generally weak but statistically significant for all criteria except DPS crashes, which were not significant. Conscientiousness was not significantly related to any of the criteria in the predictive design. In contrast, with the exception of DPS crashes and tickets, all correlations for speed were significant in the predictive design. The only significant speed correlation in the postdictive design was for self-report tickets. Combining the two criteria (self-report and DPS) did not result in validities stronger than that obtained for self-report criteria alone.

TABLE 1: Descriptive Statistics for Study Variables and Correlations and Standardized Mean Differences Between Self-Report and Archival Crashes and Moving Violation Data

Variable	Mean	SD	Sum	Min	Max	r	d
			Time 1 Data				
Conscientiousness	133.47	17.89	—	78.00	174.00		
Agreeableness	139.00	15.57	—	81.00	177.00		
Emotional stability	101.12	20.38	—	41.00	156.00		
Extraversion	114.48	20.89	—	54.00	166.00		
Openness	129.88	19.75	—	76.00	178.00		
Driving speed	3.60	3.56	—	–15.00	25.00		
Self-report defensive driving	1.41	2.32	472.00	0.00	20.00	.02	1.07*
DPS defensive driving	0.05	0.23	19.00	0.00	1.00		
Self-report crashes	0.20	0.50	68.00	0.00	3.00	.43*	0.34*
DPS crashes	0.07	0.26	22.00	0.00	2.00		
Self-report tickets	0.28	0.64	95.00	0.00	4.00	.37*	0.49*
DPS tickets	0.06	0.26	19.00	0.00	2.00		
			Time 2 data				
Self-report crashes	0.22	0.54	74.00	0.00	3.00	.41*	0.54*
DPS crashes	0.03	0.16	9.00	0.00	1.00		
Self-report tickets	0.38	0.69	126.00	0.00	4.00	.22*	0.67*
DPS tickets	0.06	0.26	20.00	0.00	2.00		

Note. Driving speed = typical average driving speed in miles per hour above posted speed limits. Defensive driving = number of defensive driving courses taken. DPS = Department of Public Safety (archival data). Sums are not reported where they are not inherently meaningful in the context of specified variable. The *d*s represent standardized difference between self-report and archival data means; significance tests are for differences between the means.

*$p < .01$. All tests are two-tailed.

The correlations shown in Table 3 also reflect poor stability in predictor/criterion relationships across the postdictive and predictive designs. Furthermore, neither the postdictive nor predictive validities were consistently stronger than the other. Statistically significant postdictive validities were generally obtained for conscientiousness, but the predictive validities for con-

TABLE 2: Cross-Classifications of Participants by Self-Report and DPS-Recorded Data

		Self-Report	
DPS		No	Yes
No	Time 1 crashes	273 (82)	40 (12)
	Time 2 crashes	274 (82)	51 (15)
	Time 1 tickets	262 (78)	55 (16)
	Time 2 tickets	236 (71)	80 (24)
Yes	Time 1 crashes	5 (2)	16 (5)
	Time 2 crashes	1 (0)	8 (2)
	Time 1 tickets	3 (1)	14 (4)
	Time 2 tickets	4 (1)	14 (4)

Note. No = no incidents reported. Yes = one or more incident reported. Percentages are in parentheses. All chi-square tests are significant ($p < .05$).

scientiousness were nonsignificant. However, speed generally yielded statistically significant predictive validities but nonsignificant postdictive validities.

Concerning the other FFM dimensions, there was no consistent, clear pattern of results. However, agreeableness was related to Time 1 self-report tickets, a finding also reported by Arthur and Graziano (1996); none of the correlations between agreeableness and Time 2 data were significant. A reverse pattern of effects was obtained for openness, in which the correlations with Time 2 tickets were significant but none of the Time 1 correlations was significant. For emotional stability, significant correlations were obtained for Time 2 self-report and DPS tickets, along with Time 1 crashes. Finally, for extraversion, none of the Time 2 correlations was significant. However, extraversion was related to Time 1 self-report crashes and DPS tickets.

DISCUSSION

The first objective of the present study was to replicate Arthur et al.'s (2001) findings concerning

TABLE 3: Correlations Between Predictors and Self-Report and Archival Data

	Predictors											
	Conscientiousness		Agreeableness		Emotional Stability		Extraversion		Openness		Driving Speed	
	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre
Self-report crashes	-.12*	.03	-.10	.03	-.02	-.07	.11*	.04	.06	.02	.10	.13*
DPS crashes	-.10	.01	-.05	-.03	.11*	-.05	.10	-.03	-.05	.05	.02	.08
Combined crashes	-.13*	.03	-.09	.02	.03	-.07	.13*	.03	.02	.03	.08	.14*
Self-report tickets	-.11*	-.10	-.11*	.04	-.04	-.18**	.10	.08	.02	.16**	.12*	.29***
DPS tickets	-.13*	-.03	-.08	.02	.03	-.12*	.11*	.02	.00	.12*	-.05	.10
Combined tickets	-.13*	-.10	-.12*	.05	-.02	-.20***	.12*	.07	.02	.18***	.08	.29***

Note. Post = postdictive correlations (i.e., criterion data were collected at Time 1). Pre = predictive correlations (i.e., criterion data were collected at Time 2). DPS = Department of Public Safety (archival data). Combined = sum of self-report and DPS data.

*p < .05, **p < .01, ***p < .001. All tests are two-tailed.

the lack of convergence between self-report and archival data sources of motor vehicle crash involvement and moving violations using a 2-year longitudinal follow-up data set. Because their study used a postdictive design, and given the potential methodological problems associated with these designs, the second objective of the present study was to investigate the stability of Arthur et al.'s (2001) predictor/criterion relationships using a 2-year longitudinal predictive design. Our results replicated the findings of Arthur et al. (2001). Specifically, the convergence between self-reports and archival records of crashes and tickets was low. Furthermore, the degree of agreement between the two sources was lower for the predictive design than it was for the postdictive design. For both designs, the lack of convergence was attributable to participants reporting substantially more incidents than were in their DPS records. Finally, there was not a clear pattern of predictor/criterion relationships across the two designs, and the magnitudes of the effects were generally small but in the range of those reported in previous research (e.g., Arthur et al., 1991; Arthur & Graziano, 1996).

The results of the present study have several implications for future motor vehicle crash involvement research. First, our results generally suggest that in studies of the relationship between specified variables and motor vehicle crash involvement, the choice of both criterion type and validation type may have an important effect on the conclusions drawn from the results obtained. The import of this implication is highlighted by the results of a detailed review of crash-related studies published in the May 2000 through May 2003 volumes of *Accident Analysis and Prevention, Ergonomics, Human Factors, Journal of Safety Research,* and *Transportation Research, Part F: Traffic Psychology and Behavior.* Of the 45 articles identified, 32 used a postdictive design (29 with archival data, 12 with self-report data, and 1 with a combination of archival and self-report data) and 3 used a predictive design (1 with archival data, 1 with self-report data, and 1 with a combination of archival and self-report data).

Consequently, given the results of the present study, one wonders whether different predictor/criterion relationships would characterize the

field if predictive designs were more predominant. Specifically, if predictive designs were more prevalent, would different conclusions be made regarding the role that specified individual difference variables play in driving behavior? For example, previous conclusions in the extant literature regarding the extent to which conscientious individuals are involved in fewer car crashes have been based exclusively on postdictive designs demonstrating a negative relationship between conscientiousness and crash involvement. However, because this pattern of results was not obtained with predictive designs, it leads one to wonder whether the conscientiousness/crash involvement relationship reported in the extant literature may not be an artifact of the postdictive design. Research needs to explore the relationship between individual difference variables and driving behavior using predictive designs.

Second, the present study suggests that self-report and archival data may represent different facets of the criterion space. In the case of archival data, crash reports may vary as a function of reporting thresholds, interstate information exchange, police investigation policies, and record-keeping policies and practices. Given these boundary conditions, one can easily envision circumstances in which self-reports may be more advantageous than archival data. Thus the larger number of incidents obtained via self-reports could be attributable to the threshold used to separate crashes ("accidents") from "minor fender benders" or "just a scratch," with self-reports having a lower threshold than that of state-reported archival data. So, self-reports potentially have the ability to canvass a broader range of incidents. Although the results show that self-reports are not inherently inferior to archival data (Arthur et al., 2001), they do not provide any information as to which is more accurate, and future research still needs to investigate this issue. Ultimately, it seems that the choice between self-reports and archival data may depend on the purpose of the investigation. For instance, if one is interested in canvassing a wider range of crashes, including minor incidents and fender benders, then self-reports may be more appropriate than archival data. However, if one is interested only in more serious crashes, especially those that may have involved

fatalities, then archival data may be more appropriate.

Given the problems associated with the collection of both predictive and postdictive "real-world" crash data, researchers can turn to performance simulations (e.g., Cnossen, Rothengatter, & Meijman, 2000; Haigney, Taylor, & Westerman, 2000; Ranney & Pulling, 1989, 1990). For example, Ranney and Pulling (1989, 1990) used a driving performance work sample consisting of 30-min "trips," with each trip composed of 20 laps on a closed course. The driver was required to respond to a continuous sequence of situations that included tasks such as responding to traffic signals and signs and selecting routes. Although they are not as frequently used in the individual differences/crash research, simulations have several advantages over archival or self-report crash data. These include the ability to compress time and permit a finer operationalization and measurement of the criterion; the ability to manipulate and control the driving environment; and being conducive to the use of concurrent criterion-related validation designs.

In summary, the present study used a predictive criterion-related validation design to replicate Arthur et al.'s (2001) results, which indicated a lack of convergence for archival and self-report data using a postdictive design. Specifically, higher numbers of crashes and tickets were obtained with self-report data than with DPS archival data. The stability of predictor/criteria relationships across criterion-related validation design (i.e., predictive and postdictive) was also assessed. Driving speed emerged as a predictor of both self-report tickets and self-report crashes in the predictive design. In contrast, whereas conscientiousness was related to self-reported crashes and tickets in the postdictive design, these relationships were not replicated in the predictive design. Thus the results from the present study suggest that the choice of validation design (i.e., predictive and postdictive) and the choice of data source (i.e., archival and self-report) are potentially important considerations in studies investigating the relationship between individual difference variables and motor vehicle crash involvement.

REFERENCES

Arthur, W., Jr., Barrett, G. V., & Alexander, R. A. (1991). Prediction of vehicular accident involvement: A meta-analysis. *Human Performance, 4,* 89–105.

Arthur, W., Jr., Barrett, G. V., & Doverspike, D. (1990). Validation of an information-processing based test battery for the prediction of handling accidents among petroleum-product transport drivers. *Journal of Applied Psychology, 75,* 621–628.

Arthur, W., Jr., & Doverspike, D. (1992). Locus of control and auditory selective attention as predictors of driving accident involvement: A comparative longitudinal investigation. *Journal of Safety Research, 23,* 73–80.

Arthur, W., Jr., & Doverspike, D. (2001). Predicting motor vehicle crash involvement from a personality measure and a driving knowledge test. *Journal of Prevention and Intervention in the Community, 22,* 35–42.

Arthur, W., Jr., & Graziano, W. G. (1996). The five-factor model, conscientiousness, and driving accident involvement. *Journal of Personality, 64,* 593–618.

Arthur, W., Jr., Tubre, T. C., Day, E., Sheehan, M. K., Sanchez-Ku, M. L., Paul, D. S., et al. (2001). Motor vehicle crash involvement and moving violations: Convergence of self-report and archival data. *Human Factors, 43,* 1–11.

Ball, K., & Owsley, C. (1991). Identifying correlates of accident involvement for the older driver. *Human Factors, 33,* 583–595.

Burns, P. C., & Wilde, G. J. S. (1995). Risk taking in male taxi drivers: Relationships among personality, observational data and driver records. *Personality and Individual Differences, 18,* 267–278.

Cnossen, F., Rothengatter, T., & Meijman, T. (2000). Strategic changes in task performance in simulated car driving as an adaptive response to task demands. *Transportation Research Part F: Traffic Psychology and Behaviour, 3,* 123–140.

Cohen, J. (1992). A power primer. *American Psychologist, 112,* 155–159.

Dalziel, J. R., & Job, R. F. S. (1997). Motor vehicle accidents, fatigue, and optimism bias in taxi drivers. *Accident Analysis and Prevention, 29,* 489–494.

Digman, J. M., & Inouye, J. (1986). Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology, 50,* 116–123.

Eden, D. (2002). Replication, meta-analysis, scientific progress, and AMJ's publication policy. *Academy of Management Journal, 45,* 841–846.

Elander, J., West, R., & French, D. (1993). Behavioral correlates of individual differences in road-traffic crash risk: An examination of methods and findings. *Psychological Bulletin, 113,* 279–294.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4,* 26–42.

Graziano, W. G., & Ward, D. (1992). Probing the Big Five in adolescence: Personality and adjustment during a developmental transition. *Journal of Personality, 60,* 425–439.

Guastello, S. J. (1993). Do we really know how well our occupational accident prevention programs work? *Safety Science, 16,* 445–463.

Haigney, D. E., Taylor, R. G., & Westerman, S. J. (2000). Concurrent mobile (cellular) phone use and driving performance: Task demand characteristics and compensatory processes. *Transportation Research Part F: Traffic Psychology and Behaviour, 3,* 113–121.

Hansen, C. P. (1989). A causal model of the relationship among accidents, biodata, personality, and cognitive factors. *Journal of Applied Psychology, 74,* 81–90.

Harano, R. M., Peck, R. C., & McBride, R. S. (1975). The prediction of accident liability through biographical data and psychometric tests. *Journal of Safety Research, 7,* 16–52.

Hendrick, C. (1990). Replications, strict replications, and conceptual replications: Are they important? *Journal of Social Behavior and Personality, 5*(4), 41–50.

Hilakivi, I., Veilahti, J., Asplund, P., Sinivuo, J., Laitinen, L., & Koskenvuo, K. (1989). A sixteen-factor personality test for predicting automobile driving accidents of young drivers. *Accident Analysis and Prevention, 21,* 413–418.

Hogan, R. T. (1983). A socioanalytic theory of personality. In M. M. Page (Ed.), *Nebraska Symposium on Motivation* (Vol. 31, pp. 55–89). Lincoln, NE: University of Nebraska Press.

Horswill, M. S., & McKenna, F. P. (1999). The development, validation, and application of a video-based technique for measuring an everyday risk-taking behavior: Drivers' speed choice. *Journal of Applied Psychology, 84,* 977–985.

Kelly, C. W., Chase, L. J., & Tucker, R. K. (1979). Replication in experimental communication research: An analysis. *Human Communication Research, 5,* 338–342.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48,* 518–537.

Marottoli, R. A., Cooney, L. M., & Tinetti, M. E. (1997). Self-report versus state records for identifying crashes among older drivers. *Journal of Gerontology, 52A,* M184–M187.

McGuire, F. L. (1973). The nature of bias in official accident and violation records. *Journal of Applied Psychology, 57,* 300–305.

McGwin, G., Jr., Owsley, C., & Ball, K. (1998). Identifying crash involvement among older drivers: Agreement between self-report and state records. *Accident Analysis and Prevention, 30,* 781–791.

National Safety Council. (2002). *Injury facts*® (2002 ed.). Itasca, IL: Author.

Owsley, C., Ball, K., Sloane, M. E., Roenker, D. L., & Bruni, J. R. (1991). Visual/cognitive correlates of vehicle accidents in older drivers. *Psychology and Aging, 6,* 403–415.

Ranney, T. A., & Pulling, N. H. (1989). Relation of individual differences in information-processing ability to driving performance. In *Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 965–969). Santa Monica, CA: Human Factors and Ergonomics Society.

Ranney, T. A., & Pulling, N. H. (1990, January). *Performance differences on driving and laboratory tasks between drivers of different ages.* Paper presented at the 69th Annual Meeting of the Transportation Research Board, Washington, DC.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54,* 93–105.

Sleek, S. (1996). Training the road warrior. *APA Monitor, 27,* 13–14.

Smith, D. I. (1976). Official driver records and self-reports as sources of accident and conviction data for research purposes. *Accident Analysis and Prevention, 8,* 207–211.

Szlyk, J. P., Alexander, K. R., Severing, K., & Fishman, G. A. (1992). Assessment of driving performance in patients with retinitis pigmentosa. *Archives of Ophthalmology, 110,* 1709–1713.

Szlyk, J. P., Fishman, G. A., Severing, K., Alexander, K. R., & Viana, M. (1993). Evaluation of driving performance in patients with juvenile macular dystrophies. *Archives of Ophthalmology, 111,* 207–212.

Szlyk, J. P., Seiple, W., & Viana, M. (1995). Relative effects of age and compromised vision on driving performance. *Human Performance, 37,* 430–436.

Tomlinson-Keasey, C., & Little, T. D. (1990). Predicting educational attainment, occupational achievement, intellectual skill, and personal adjustment among gifted men and women. *Journal of Educational Psychology, 82,* 442–455.

Wasielewski, P. (1984). Speed as a measure of driver risk: Observed speeds versus driver and vehicle characteristics. *Accident Analysis and Prevention, 16,* 89–103.

West, R., French, D., Kemp, R., & Elander, J. (1993). Direct observation of driving, self reports of driving behaviour, and accident involvement. *Ergonomics, 36,* 557–567.

Wilson, T., & Greensmith, J. (1983). Multivariate analysis of the relationship between drivometer variables and drivers' accident, sex, and exposure variables. *Human Factors, 25,* 303–312.

Zylman, R. (1972). Drivers' records: Are they a valid measure of driving behaviour? *Accident Analysis and Prevention, 4,* 333–349.

Winfred Arthur, Jr., is a professor in the Psychology Department at Texas A&M University. He received his Ph.D. in industrial/organizational psychology in 1988 from the University of Akron, Ohio.

Suzanne T. Bell is an assistant professor in the Psychology Department at DePaul University. She received her Ph.D. in industrial/organizational psychology in 2004 from Texas A&M University.

Bryan D. Edwards is an assistant professor in the Psychology Department at Tulane University. He received his Ph.D. in industrial/organizational psychology in 2003 from Texas A&M University.

Eric Anthony Day is an assistant professor in the Psychology Department at the University of Oklahoma. He received his Ph.D. in industrial/organizational psychology in 1998 from Texas A&M University.

Travis C. Tubre is an assistant professor in the Psychology Department at the University of Wisconsin-River Falls. He received his Ph.D. in industrial/organizational psychology in 2000 from Texas A&M University.

Amber H. Tubre is the outreach program manager at the University of Wisconsin-River Falls. She received her M.S. in industrial/organizational psychology in 2000 from Texas A&M University