SELF-DECEPTION AND IMPRESSION MANAGEMENT IN COMMERCIAL PILOTS: AN UNDERREPORTED AND POTENTIAL CONFOUND IN AVIATION RESEARCH

Tabitha T. Black
Eastern Washington University
Cheney, WA
Charalambos C. Cleanthous
Eastern Washington University
Cheney, WA
Ryan Sain
PsyCore
Spokane, WA
Steven M. Vera
Eastern Washington University
Cheney, WA

A concern when administering questionnaires is whether the participant is providing information that is accurate. The *Balanced Inventory of Desirable Responding* (BIDR) was used to assess commercial pilots' socially desirable responding resulting in two profiles: Impression management (IM; faking bad) and self-deceptive enhancement (SDE; faking good). These pilots' profiles were compared to the *Aviation Safety Locus of Control* (ASLOC) scale, used to measure external (ASLOC-E) or internal (ASLOC-I) orientation, and the *Crew Resource Management Training Survey* (CRMTS) developed from the Federal Aviation Administration's guidelines for CRM. The results from the SDE indicated that over a fourth of the participants responded in a socially desirable manner. Significant differences were also found between those scoring high on the IM subscale versus those scoring in the normal range of the CRMTS subscales.

The well documented issue of socially desirable responding continues to present a self-report validity concern to behavioral science researchers (van de Mortel, 2008). If uncontrolled, it may confound the validity of research results (Nederhof, 1985). Socially desirable responding occurs when a participant's response bias results in answering survey questions that present the participant in a favorable light. This bias is a function of the test behavior of the subject (King, Bruner, & Hensel, 1991) and not necessarily always indicative of malicious intent. SDR can result in self-deception, in which the participant believes the presentation to be true about oneself (Nederhof, 1985; Paulhus, 1991). Alternatively, the response bias may result in the attempt to present oneself as worse off than what one is, this is known as impression management (Paulhus, 1991). Several factors contribute to SDR including the test setting, participant motives, and the participant's expectation of repurcussions of responses (King, Bruner, & Hensel, 1991). Both qualitative (Stodel, 2015) and quantitative research efforts have attempted to identify SDR biased responding. One of the early attempts at quantifying the construct through the use of a questionnaire, that had acceptable validity and reliability was the Crowne-Marlow scale (Crowne & Marlowe, 1960). A variety of methods have been devised to control SDR, yet none are efficacious in controlling the response bias for specific settings or intended populations (Hunter & Stewart, 2009; Nederhof, 1985).

There is a dearth of research regarding SDR and commercial pilots, but the research that does exist strongly suggests that there is a need for this measure when assessing pilot responses (Butcher & Han, 1995). Pilots, through the process of aviation training and testing, become well adapted to positive self-presentation for the purposes of career advancement (Butcher & Han, 1995). They spend their career in regular training and testing for the purpose of maintaining proficiency in their work demands. This training consists of both written and oral tests, and concurrent validity of maneuver performance, also known as practical tests, of maneuvers while under the supervision of an examining authority. Interweaved into the fabric of crew resource management (CRM) training, exists a cultivation of personal confidence, assertiveness and authority as pilot in command. SDR factors, setting, personal motives, and expectancies (King, 1991) when applied in aviation, may or may not be amplified (Galic, Jerneic, & Kovacic, 2012). Moreover, individuals who have a motive to present themselves in superlative manner, such as commercial pilots, consistently produce higher profiles on defensiveness indexes (Butcher & Han, 1995). The concern over pilot defensiveness has led to the development of a scale specifically applicable to commercial pilots (Butcher & Han, 1995).

Directly relevant to SDR, is CRM as the in-cockpit activities directly impact the outcome of any given flight. A co-pilot, for example, who does not speak up when one should, due to one's self-preservation concern, a potential SDR issue, directly impacts the outcome of the flight. CRM training in the USA has been in effect for approximately two decades (Helmreich, Merritt, & Wilhelm, 1999). Overall, crewmembers find CRM training to be relevant and useful, and the aviation industry within the US has overwhelmingly endorsed the program (Helmreich & Wilhelm, 1991). The foundation of the concept and the challenges it aims to address have been as a result of workshops and meetings initiated by airlines and aerospace authorities for the purpose of aviation safety for the last 40 years. The evolution of CRM has been a reactive analysis initiated by various aviation forums (Helmreich, Merritt, & Wilhelm, 1999). Although generally accepted, one continuing question in CRM is its effectiveness and validity (Salas, 2006). One approach to investigating CRM's effectiveness is through aviation psychological research, that is, observation of CRM performance and surveying pilots' opinions and attitudes. CRM method performance and inquiry of the pilot. In order to evaluate the research, one must understand the degree to which participant bias affects the resultant data.

An area in which pilots, overall, have excelled is locus of control, that is, they exhibit an internal locus of control indicating that they are responsible and capable of dealing with events (Hunter, 2002; Skinner, 2011). In the aviation environment safety is paramount; the perceived locus of control of an event is important to predicting the outcome of emergency situations (Hunter, 2002). The research indicates that pilots who are at greater risk of aircraft accidents can be identified in advance (Hunter, 2002). The aviation safety locus of control scale (ASLOC) has been useful in this regard (Hunter, 2002). Given the concern with SDR, however, the question that arises is the extent to which a pilot's SDE or IM impact the degree of perceived control, and, CRM. Given the concern with SDR and the dearth of research in the area, the question that arises is whether those scoring high on either the SDE or IM also demonstrate an external ASLOC and, consequently, making them poorer managers of emergency situations.

The present study sought to examine whether pilots who endorsed socially desirable items had significantly different profiles than those who did not. If differences were identified, if those differences led to a lower endorsement of various crew resource management criteria as promoted by USA Federal Aviation Administration guidelines (Federal Aviation Administration, 2004) or an external locus of control profile.

Methods

Participants

With the permission from the site administrators, a link to the Crew Resource Management Study was posted on the Flights Above the Pacific Northwest Facebook page (https://www.facebook.com/groups/FLightsAboveThePNW/) and Airline Pilot Central Forums (http://www.airlinepilotforums.com/). Participation of the survey was restricted to USA/FAA commercial rated pilots, employed as an active pilot within the last 10 years. This restriction ensured that all participants would have completed a formalized CRM training per FAA regulation AC 120-51E (Federal Aviation Administation, 2004). The results indicated that the distribution by participant gender matched the current ratio of employed commercial pilots within the US: 65 total participants: 58 males, 5 females, 2 gender non-response (Bureau of Labor Statistics, 2016).

Measures

The *Balanced Inventory of Desirable Responding* (BIDR) was developed to measure two dimensions of SDR (Paulhus, 1991) that were absent in prior measures. The BIDR is comprised of 40 7-point Likert-type scale items. Paulhus (1999) reported convergent validity for SDE subscale with several other scales including, among others, repressive styles, defense mechanisms, and ways of coping. Convergent validity for the IM subscale was reported Eysenck's Lie scale and the MMPI Lie scale. The internal reliability resulted in Cronbach's alpha for IM=.84 and SDE=.75. Convergent validity for the SDE subscale was reported with several other scales including, among others, repressive styles, defense mechanisms, and ways of coping. The first subscale is referred to as impression management (IM) and is a bias that reflects a person's attempt to present oneself in an unrealistically positive manner; it is also referred to as faking good. The self-deceptive enhancement (SDE) index, the second subscale, measures the behavioral response tendency of a to answer items in a manner that portrays oneself in a positive light.

The *Aviation Safety Locus of Control* (ASLOC) is unique in this domain of scales as it was designed specifically to measure safety issues within an aviation environment; two subscales are produced from the 20 items: external (ASLOC-E) and internal safety locus (ASLOC-I) (Hunter, 2002). The items are presented on a 5 point Likert-type scale (*strongly agree* to *strongly disagree*). Hunter (2002) reported that the two subscales of the ASLOC exhibited acceptable internal consistency and were negatively correlated ($r = - 0.419$, $p < 0.001$). Construct validity was reported by comparing the combined ASLOC scores with the resignation score from the Hazardous Attitudes Inventory.

A demographic survey was designed to provide a description of the participants including: gender, post-secondary education and total flight time. The *Crew Resource Management Training Survey* (CRMTS) was developed (Black) from the FAA guidelines (Federal Aviation Administration, 2004) for CRM training to assess the participants and opinions and self-reported use of CRM training procedures. The CRMTS was comprised of seven subscales including: pilot in command (PIC), communication (COM), management of a flight team (MFT), time management (TM), fatigue (FTG), stress (STR), and aeronautical decision making (ADM). The eighty items were made up of three response styles based on content: 5 point Likert-type scale (*strongly agree* to *strongly disagree*), true-false, and multiple choice items.

Results

There were 68 anonymous participants. Three were dropped from the analysis as their data was too incomplete to use. Of the 65 remaining participants, 5 were female and 58 males. These numbers match the current female-male ratio of pilots in the US. The small number of female participants makes it impossible to conduct adequate statistical analyses and will only be used to indicate trends. In terms of flight time, 8 (7 male, 1 female) pilots had between 1000-2000 hours, 16 (13 male, 3 female) between 2000-4000 hours, 17 (16 male, 1 female) between 5000-10,000 hours, and 22 (all male) had over 10,000 hours. The participants' educational level, inclusive of either completed or earned degree, 8 had an AA or AS, 33 a BA or BS, 21 with an MA or MS, and 2 with a doctorate.

The overall results showed that 48 pilots scored in the normal range on the BIDR, 13 scored high on IM and 4 on SDE. Table 1 shows the results from a correlational analysis of the relationship between the BIDR and ASLOC scales. Table 1 shows the significant correlations between the BIDR, ASLOC, and CRM. The only correlation that was significant between the BIDR and ASLOC was between the SDE and ASLOC-E that are positively correlated. There was a significant negative correlation between the ASLOC-I and time management. The ASLOC-E was also positively correlated with time management, and negatively correlated with pilot-in-command.

Table 1.

Pearson Product Moment Correlation: BIDR and ASLOC Subscales

	N	r	Sig.
BIDR (SDE) x ASLOC-E	64	.25	.047
ASLOC-I x CRM--Time Management	64	-.31	.012
ASLOC-E x CRM--Time Management	64	.34	.006
ASLOC-E x CRM--Pilot-in-Command	63	-.35	.005

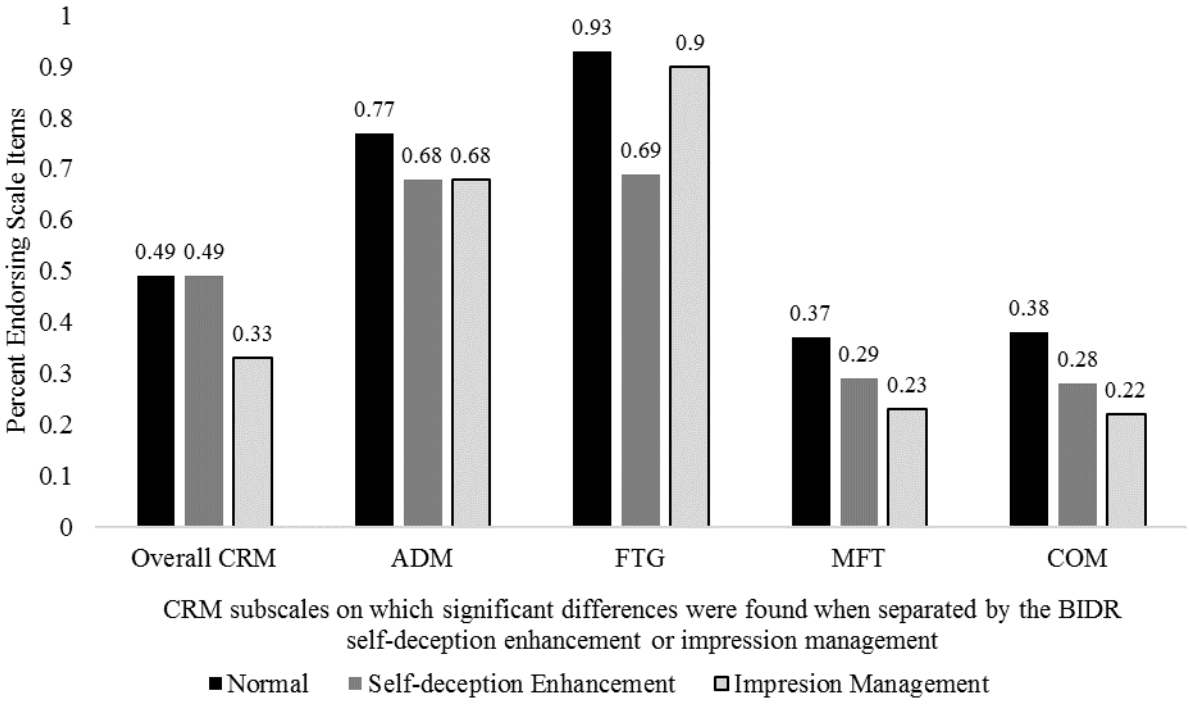
The comparison of the BIDR IM and SDE and CRM subscales indicated several important results. Significant differences were found on four of the seven subscales and the overall CRM score. Comparisons across all groups were made using the Kruskal-Wallis independents samples test. Overall CRM, ADM, FTG, MFT, and COM subscales were significant at the .003, .006, .002, .029, and .009 levels, respectively. All follow-up comparisons using a Mann-Whitney were not significant after a Bonferroni correction or were invalid. Figure 1

shows that where there were significant differences, the normal group endorsed more CRM items than either the SDE or the IM group.

The overall CRM score indicated that the normal and the SDE group endorsed an average of 49% of the items, whereas the IM group endorsed 33%. On the ADM subscale, the normal group endorsed 77% and the SDE and IM group both endorsed 68% of the items. On the FTG subscale, the normal group endorsed 93% whereas the SDE group endorsed 90% and the IM group 69% of the items; this was the only subscale on which the SDE group endorsed more items than the IM group. The MFT subscale indicated that the normal group endorsed 37%, the SDE 29%, and the IM group endorsed 23% of the items. Finally, the COM results indicated that the normal group endorsed 38%, the SDE group 28%, and the IM 22% of the items.

Figure 1.

Comparison of BIDR on CRM Endorsement



Discussion

The results speak to the necessity of using a scale that assesses socially desirable responding in aviation research. Although a variety of procedures have been developed to counteract socially desirable responding, all leave much room for improvement. In the present case, rather than excluding the socially desirable responders, their results made up the comparison group allowing us to detect important differences when compared to the typical responders.

The positive correlation between SDE and ASLOC-E, although relatively small, does suggest that these two factors are influencing one another. However, given that it is a weak relationship, it is possible that extreme scores are impacting this relationship. Further examination of this relationship is warranted, especially as the correlation is between an external locus of control and someone who is faking good. In the ASLOC-E case, it is someone who may be attributing to factors out of one's control, not the best strategy in the cockpit, and someone who is exaggerating one's strengths. The latter is dangerous as it may be masking one's limitations (King, 1991). Two correlations in opposite directions were found depending on whether the relationship was between ASLOC-I or ASLOC-E and TM. One group is indicating that TM is a problem (ASLOC-E) and the other (ASLOC-I) is suggesting that time management is not such a big problem. Further investigation of how these attitudes are

impacting one's in-cockpit and work-related behavior is important as time is inescapable. The final significant correlation was between ASLOC-E and PIC. The results indicate that those scoring high on ASLOC-E had a diminishing view of the PIC's role. Again, given the nature of the cockpit and the necessity of working as a team, further exploration of this issue is warranted.

The other analyses were concerned with SDR and CRM. Examining the overall CRM endorsement, two conclusions can be reached. First, the overall endorsement of CRM practices is under 50%. Second, the IM group endorses CRM practices at 33%. The overall results suggest that there is much room for improvement irrespective of ones SDE or IM score. However, a finer grained analysis indicates that there is considerable variation in the items endorsed by all groups. Such a distribution argues for an interpretation at the individual scale level. By examining the items that are endorsed or not endorsed, would allow for further refinement of both the scale and more targeted CRM interventions. In other words, training would be targeted for those weak in the area of CRM. Based on that assessment of the individual subscales, one would have various groups to target, that is, the groups would be made up of the subscales on which one was weak, potentially, that could be all seven subscales. An important use of the overall score could be to determine weaknesses in an individual's understanding of CRM. Having identified such an understanding a more targeted intervention could be developed. All this can take place before the pilot, for obvious CRM reasons, is allowed in the cockpit. Moreover, such a targeted intervention could be used as a continual assessment of the impact of CRM training.

There were two limitations to the present study. The first limitation was the development of the CRMTS. It is a rationally developed survey based on the criteria established by the FAA. It is crucially important that such a survey be developed with the appropriate psychometric properties. The second limitation concerned the participant sample. In particular, the concern is with the low number of female participants. Even though the percentage of participants matched the USA commercial pilot rates, the low number made it impossible to draw any statistically meaningful conclusions. Obviously, it behooves researchers to pursue this matter with some urgency as the females displayed a higher rate of SDR than did their male counterparts.

A strength of the present research was to use a computerized version of SDR that research indicates is the best, current, method for decreasing SDR responding. It is possible that the current rate of 27%, bad as that may be, is lower than would have been the case had paper-and-pencil assessments been used. Given the weakness and strengths of the present study, the conclusion that SDR impacts what commercial pilots are endorsing about CRM training and practices cannot be overstated. Incorporating measures to assess and control SDR responding in commercial pilots is warranted.

References

Bureau of Labor Statistics (2016). *Labor Force Statistics from the Current Population Survey.* Transportation and material moving—aircraft pilot and flight engineers. Retrieved on February 20, 2017, from: https://www.bls.gov/cps/cpsaat11

Butcher, J. N., & Han, K. (1995). Development of an MMPI-2 scale to assess the presentation of self in a superlative manner: The S Scale. In J. N. Butcher, C. D. Spielberger, J. N. Butcher, C. D. Spielberger (Eds.), *Advances in personality assessment* (vol. 10, pp. 25-50). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Crowne, D. P., & Marlowe, D. (1960). A new social scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349-354.

Flights Above The Pacific Northwest (2016). https://www.facebook.com/groups/FLightsAboveThePNW/

Federal Aviation Administration. (2004). Advisory Circular System 120-51E (2004) – Crew resource management training. Retrieved on August 4, 2016 from http://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentid/22879

Galic, Z., Jerneic, Z., & Kovacic, M. P. (2012). Do applicants fake their personality questionnaire responses and how successful are their attempts? A case of military pilot cadet selection. *International Journal of Selection and Assessment*, 20(2), 229-242.

Helmreich, R. L., & Wilhelm, J. A., (1991). Outcomes of crew resource management training. *International Journal of Aviation Psychology, 1*(4), 287-300.

Helmreich, R. L., Merritt, A. C., & Wilhelm, J.A. (1999). The evolution of crew resource management training in commercial aviation. *International Journal of Aviation Psychology*, 9(1), 19-32.

Hunter, D. R. (2002). Development of an aviation safety locus of control scale. *Aviation, Space, and Environmental Medicine,* 73(12), 1184-1188.

Hunter, D. R., Stewart, J. E., (2009). Locus of control risk orientation, and decision making among U.S. army aviators. Technical Report 1260. Personnel, Performance and Training Technology Army Project Number 622785A790.

King, M. F., Bruner, G. C., & Hensel, P. J. A., (1991). Social desirability bias: A neglected aspect of validity testing. *Psychology and Marketing 2000, 17*(2), 79-103. Retrieved January 15, 2015 from https://www.researchgate.net/profile/Gordon_Bruner/publication/230034577

van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, 25*(4), 40-48.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*(3), 263-280.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). New York: Academic Press.

Salas, E (2006). Does crew resource management work? An update, an extension and some critical needs. *Human Factors and Ergonomics Society, 48*(2) 392-412.

Skinner, E. A., & Zimmer-Gembeck, M. J. (2011). Perceived control and the development of coping. In S. Folkman, S. Folkman (Eds.), *The Oxford handbook of stress, health, and coping* (pp. 35-59). New York, NY, US: Oxford University Press.

Stodel, M. (2015). But what will people think? Getting beyond social desirability bias by increasing cognitive load. *International Journal of Market Research, 57*(2), 313-323. DOI:10.25/IJMR-2015-024