! ""! # $$ %% &'((() *)+ ' '$,%

- + . * & $ +$ & + *
/ $ 0 1$ + + # + &
0 + 1 * * 2$ 3
4 0 *

5 6 -) 7 8 (&

2 + $ "#$% &' ()*+)*,)-)$./0123 &#456$7 8*9)$:;<*- 5=) >7;4=6:6-<4#:
#*9 ?)=#@<6$#: A=#$#45)$<75<47 6B CBB<4) 6B D5$#5)-<4 D)$@<4)7 A#*9<9#5)7E F=)
G#5# H)@<7<5)9I J6K$*#: 6B >)$76*#:<5; L77)77M)*5I NOE1I 100!110I GCPE
10'10Q0R00//SQN1'/01T'NS2NQ0

2 6 $ https://doi.org/10.1080/00223891.2014.935980

Factors Underlying the Psychological and Behavioral Characteristics of Office of Strategic Services Candidates: The Assessment of Men Data Revisited

MARK F. LENZENWEGER[1,2]

[1]Department of Psychology, The State University of New York at Binghamton
[2]Department of Psychiatry, Weill Cornell Medical College, Cornell University

During World War II, the Office of Strategic Services (OSS), the forerunner of the Central Intelligence Agency, sought the assistance of clinical psychologists and psychiatrists to establish an assessment program for evaluating candidates for the OSS. The assessment team developed a novel and rigorous program to evaluate OSS candidates. It is described in

In response to the need for professional assistance in developing a more thorough psychological and behavioral assessment adjunct to selection, the OSS reached out to a number of prominent clinical psychologists and psychiatrists in the academic community within the United States (MacKinnon, 1974/1980; OSS Assessment Staff, 1948). This established the connection between clinical psychology, with its focus on developing assessment approaches, and the nascent U.S. intelligence community. There had been no comparable prior instance (prior to World War II) in the Þelds of clinical psychology, personnel psychology, or clinical psychiatry Þelds where intensive study of individuals was carried out for the stated purpose of selection for likely suitability as an intelligence ofÞcer or special operations personnel (Banks, 2006; Williams, Picano, Roland, & Bartone, 2012; see also Butcher, 2010). In contrast to this situation in the United States, both British and German psychologists and psychiatrists had been active in assisting in the selection of ofÞcers for the military during World War I and prior to World War II (Banks, 1995). In fact, the OSS received input in 1943 regarding the nature of the British War OfÞce Selection Boards, and this served as an impetus for the OSS to engage with U.S. psychologists and psychiatrists (MacKinnon, 1974/1980).

Many well-known psychologists and psychiatrists (see Handler, 2001; OSS Assessment Staff, 1948) participated in the development of the assessment protocol that was used in the OSS assessment program. The lead was taken by Henry A.

S, the main assessment center (there were several others), and more than 5,000 candidates were assessed there in a period of about 20 months across all OSS assessment sites (Handler, 2001).[1]

The OSS assessment program at Station S and the other assessment stations clearly produced an enormous amount of empirical psychological data. However, most of these data were never analyzed fully in any detailed manner. Rudimentary analyses were presented in the Appendices of the Assessment of Men (OSS Assessment Staff, 1948), with some secondary analyses examining performance outcomes (available after the war) being conducted some years later by personality psychologist Jerry Wiggins (1973). The multivariate array of the psychological data generated by the OSS assessments, which likely harbors interesting factors, was subject to a limited multivariate analysis right after the war, namely an exploratory factor analysis (EFA). That factor analysis sought to reduce the large number of assessment variables to a smaller set of underlying factors. This early analysis, carried out by the OSS Assessment team and reported in Assessment of Men (OSS Assessment Staff, 1948), was done essentially by hand and, as such, was not as richly or precisely conducted as can be done with modern methods. The factor analysis solution reported in the Assessment of Men was entirely exploratory in nature. Moreover, the methods at the time did not allow the original psychological investigators to determine which of several competing substantive models might provide the best fit to their data. This was so because computational technology was limited in the late 1940s, which limited the form of EFAs that could be done. Also, importantly, the statistical approach known as confirmatory factor analysis (CFA) simply did not exist at that time.

This study, therefore, seeks to apply modern statistical methods to the analysis of the multivariate data available from the original OSS assessments done at Station S for a subset of candidates evaluated there. The data available for this study are contained in a correlation matrix relating the

METHOD

Subjects

The subjects for this study were 133 candidates from several of the final OSS candidate classes evaluated at Station S. The data from these 133 subjects were presented in the form of a published correlation matrix in the Appendix (OSS Assessment Staff, 1948, Appendix B, p. 510). The matrix is based on complete data on all 133 subjects for all measures.

Data Structure

The battery of measures, assessment devices, behavioral task situations, and interviews (including stress interviews) used to assess OSS candidates was extensive and described in extensive detail in Assessment of Men (OSS Assessment Staff, 1948). This massive corpus of assessment data was used as the basis for the clinical ratings on 11 core dimensions for each candidate by the original assessment staff as described earlier. The data analyzed for this study is a correlation matrix (Table 1) relating these 11 variables (with unities [1.00] placed initially in the diagonal). The original correlation matrix contained 11 psychological, personality, or behavioral variables (the 10 variables plus an “overall” rating; see later). The dimensions are: (a) motivation for the assignment,[2] (b) energy and initiative, (c) practical (effective) intelligence, (d) emotional stability, (e) social relationships, (f) leadership, (g) security (i.e., ablilty to keep secrets, ability to bluff, maintain cover), (h) physical ability, (i) observing and reporting, and (j) propaganda skills. Each of these corresponded to a quantitative dimension on which the candidates were evaluated, and summary ratings were made by the OSS assessment staff. An important concern regarding the EFA done originally by the OSS assessment staff was that it used the correlation matrix described here, but it also included an additional (11th) variable. The additional variable was described as an “overall” summary rating. This overall rating was clearly highly redundant with the other variables reported because it was statistically infused with the ratings of the other 10 variables under consideration. Inclusion of the overall rating variable in the original matrix probably introduced a statistical artifact into the original analysis, which probably interfered with model estimation and could have contributed to error inflation in the factor analysis. Moreover, perhaps more important, the exact meaning of the overall rating was clearly elusive even to the OSS assessment staff, who described it as “an estimate of the total potentialities of the candidates for meeting the challenges of life—exceedingly vague and difficult concept to define” (OSS Assessment Staff, 1948, p. 217). For the purposes of this analysis, the overall variable has been eliminated. The subject are then statistically contrasted to determine which of those models tested fits the data best (see Lenzenweger, Dworkin, & Wethington, 1989, for an extensive discussion of the merits of CFA over EFA approaches). The CFA approach employed here made use of maximum-likelihood estimation for computations and several well-established indexes or procedures for the evaluation of the fit between stand-alone theoretical models and the observed results (goodness of fit chi-square, Akaike Information Criterion [AIC], Comparative Fit Index [CFI], standardized root mean square residual [SRMR]). Competing models were compared to one another using the chi-square difference test as well as the Tucker—Lewis incremental fit index. The latter evaluates improvement in fit for a model of interest as contrasted with a null model. The LISREL 8.0 program (Version 8.80, Jöreskog & Sörbom, 2006) was used to conduct the CFAs.

Primary Competing Models Estimated With OSS Data

CFAs were carried out in a stepwise manner in which nested models were systematically evaluated for their fit to the data and their relative fit with respect to each other, taken in succession. Formulation of these models was influenced in part by the EFA results as well as consideration of the psychological and behavioral features under study. Models involving two or three factors allowed the latent factors to be correlated. The models estimated are detailed as follows:

1. A null model (where all model parameters were fixed) was estimated that assumed no common latent structure. Although not truly plausible, the null model provides a good baseline against models that do make explicit assumptions regarding latent structure. (A null model is estimated to determine whether or not it can be rejected. There would be no point in modeling a data set in which all variables were uncorrelated.)
2. A one-factor model that assumed all features loaded on a single common underlying factor. Such a model is reasonable, as the assessment staff were taking a whole person, holistic approach, which might have yielded a highly interrelated set of variables in the final assessments.
3. A two-factor model was formulated that partitioned interpersonal or social and emotional variables (social relations, emotional stability, motivation, energy and initiative, leadership, physical ability, security) from the intelligence processing (effective IQ, propaganda skills, observing and reporting) variables. Such a model allows interpersonal or social features to be linked with emotional stability, holding aside processes clearly linked to cognitive and information processing capabilities. Security is linked to the emotional factors, as it is assumed that one's capacity to

rotation. A solution with three factors was retained based on both the Kaiser criterion (eigenvalue ≥ 1.0) and the Scree test (Gorsuch, 1983). Highly similar results to those reported in Table 2 were obtained for the EFA whether using principal axis factoring or, alternatively, principal components analysis. Moreover, the results (number of factors, patterning of loadings) were largely the same whether using an orthogonal (Varimax) or oblique (Oblimin) rotation. What can be seen from Table 2 is that the solution retained is somewhat similar (but clearly not identical) in pattern to the original centroid solution retained by the OSS Assessment staff (see Table 3), but some important differences appear between the two solutions. For example, in the new EFA, Factor 1 consists of effective IQ, propaganda skills, and observing and reporting, but also contains a substantial loading on leadership (which was not present in the original analysis). Factor 2 (Table 2) consists of emotional and interpersonal adjustment items (emotional stability, social relations, security, motivation for assignment), whereas in the original EFA (Table 3) the "adjustment" factor did not load the variable motivation for assignment heavily. Finally, Factor 3 (Table 2) appears to be a factor accounting for agentic or surgent behaviors, consisting of energy and initiative, physical ability, and leadership. It is interesting that the motivation for assignment variable had something of a weak relationship with all of the factors obtained in the original OSS EFA (Table 3), but loads Factor 2 substantially in this analysis.

Interpretation of the original EFA results (Table 3) is hindered by the fact that, as noted earlier, the details of the original factor extraction method (centroid method) and factor rotation method (if any) were not specified in the original report (OSS Assessment Staff, 1948). This information would be considered critical to understanding the results of a factor analysis as reported by today's standards. This missing information is important given some of the striking differences between the modern EFA and the 1948 EFA. For example, as can be seen in the original solution (Table 3), four factors were retained from the analysis of the correlation matrix reported in 1948 and were interpreted by the OSS staff, but only three were retained in the current EFA. It is important to note that even when the complete matrix is analyzed (i.e., including the "overall" variable) using modern software, the results do not support retention of a fourth factor (the fourth factor in the current analysis has an eigenvalue of .77, or well short of the 1.00 that would be needed customarily to justify retention). It is also particularly striking that the factor loadings computed with modern statistical software differ considerably from those reported in the original 1948 analysis (compare loadings in Table 2 with Table 3). That said, the general similarity of the pattern040sp26(appea)cow0983 4 s2x(if)-475.6(a

TABLE 4.ÑFactor loadings for competing models obtained using conÞrmatory factor analysis.

	Competing Models					
	One Factor	Two Factor		Three Factor		
	1	1	2	1	2	3
OSS Variable	Unifactorial	Emotional/Interpersonal	Intelligence Processing	Intelligence Processing	Emotional/Interpersonal	Agency/Surgency
Effective IQ	.73	Ñ	.94	.99	Ñ	Ñ
Propaganda skills	.62	Ñ	.75	.71	Ñ	Ñ
Observing and reporting	.51	Ñ	.67	.64	Ñ	Ñ
Social relations	.56	.58	Ñ	Ñ	.71	Ñ
Emotional stability	.59	.65	Ñ	Ñ	.89	Ñ
Security	.29	.30	Ñ	Ñ	.42	Ñ
Motivation for assignment	.56	.58	Ñ	Ñ	Ñ	.55
Energy & initiative	.79	.84	Ñ	Ñ	Ñ	.84
Leadership	.84	.83	Ñ	Ñ	Ñ	.85
Physical ability	.31	.39	Ñ	Ñ	Ñ	.35

Note. $n = 133$. Ñ = a LISREL constrained zero loading. These solutions are direct and unique with no rotation necessary. The LISREL program allows one to estimate the degree to which the latent variables underlying the OfÞce for Strategic Services assessment dimensions are correlated in the models containing more than one latent variable (i.e., factor). For the two-factor model the correlation between emotional/interpersonal and intelligence processing = .66 ($p < .001$). For the three-factor model, the correlations were as follows: Intelligence Processing × Emotional/Interpersonal = .29 ($p < .01$); Intelligence Processing × Agency/Surgency = .69 ($p < .001$); and Emotional/Interpersonal × Agency/Surgency = .70 ($p <$

larger the chi-square value (smaller p values), the poorer the Þt between the model and data; the smaller the chi-square (large p values), the better the Þt. Inspection of the chi-square values for the four models (Table 5; null through three-factor) shows a steady decline in magnitude of the chi-square value, suggesting increasingly better Þt between the model and OSS data as one moves toward the three-factor model. Similarly, when interpreting both the AIC and SRMR as indexes of Þt, the principle is the smaller the AIC and SRMR values, the better the Þt. Inspection of the AIC and SRMR values for the four models reveals the three-factor model with the smallest AIC and SRMR values. The SRMR value for the three-factor model (.08) is suggestive of a good Þt between a model and the observed data (Hu & Bentler, 1999). The CFI, which contrasts the Þt of the model of interest with that provided by the null model, shows that three-factor model provides a generally good Þt to the data (CFI = .93), where larger values of the CFI indicate better Þt (CFI = 1.00 would indicate a perfect Þt). Finally, as noted also in Table 4, LISREL allows one to estimate the degree to which the latent variables (i.e., factors) in the two-factor and three-factor models are correlated. In short, the latent variables in both the two-factor and three-factor model are substantially correlated, however the Intelligence Processing × Emotional/Interpersonal factors are somewhat less strongly associated (although still signiÞcantly) as compared to the other factor combinations.

The next step in evaluating the CFA results is to conduct a sequential comparison of models using the differences in the goodness-of-Þt chi-square values for the four models. Thus, three contrasts were conducted: (a) the null versus one-factor model, (b) the one-factor versus two-factor model, and (c) the two-factor versus three-factor model. A comparison of the chi-square Þt statistics for each model examines the differences in the chi-square values and the degrees of freedom. The difference between these chi-square values is then evaluated for statistical signiÞcance. These differences reveal the extent to which one model Þts the data better (or worse) than a competing model. To assess the amount of information gained in the comparison of two competing models and to generate an estimate of the improvement in Þt obtained in using a better model versus the null model, the non-normed incremental Þt index (Tucker—Lewis Index [TLI]; Tucker & Lewis, 1973) was calculated. The results of the model comparisons and the cumulative incremental fit index values are in the bottom panel of Table 5. As can be seen from Table 5, the one-factor model clearly provides a significantly better fit to the data than the null model. However, the two-factor model is a significant improvement in fit over the one-factor model, and the three-factor model is a significant improvement over the two-factor model. The cumulative TLI values reveal that the three-factor model (TLI = .91) is within the realm of a good or acceptable, although not perfect, fitting model given the observed OSS data. Clearly, the stand-alone fit indexes (top panel, Table 5) and the model comparison results (bottom panel, Table 5) point to the three-factor model as providing the best fit to the observed data, with clear superiority over one-factor and two-factor models. It is worth restating that a plausible four-factor model did not fit these data well, but, in fact, generated an invalid solution. Thus, simply adding additional factors does not necessarily improve model fit. In summary, the CFA results strongly suggest that the hypothesized three-factor model described earlier provides the best fit to the 10-variable matrix generated by the original OSS ratings. These results are supportive of the current EFA results and place the proposed three-factor model on a foundation consisting of much firmer statistical information. In short, the OSS assessment team ratings of the candidates reveal three factors at play: intelligence processing, emotional or interpersonal features, and agency/surgency.

Supplementary models estimated with OSS data. In addition to these four primary CFA models, three alternate (but theoretically grounded and plausible) models were estimated in the spirit of analytic thoroughness. One was a two-factor model, the second was a three-factor model, and the final one was four-factor in nature. The alternative two-factor model was one that placed the "Security" variable with the intelligence-related (i.e., tradecraft) items. This model was considered as one might think that the ability to keep a secret might depend more closely on skills that covary with the ability to do intelligence-related activities and demonstrate good tradecraft skills (see later). This alternative two-factor model did not fit the observed data as well as the primary two-factor model that placed "Security" with the emotional or interpersonal variables (χ^2 = 158.44, AIC = 200.44, CFI = .87, SRMR = .10). Similarly, the alternative three-factor model that also kept "Security" with the intelligence-related items did not fit the observed data as well as that found for the primary three-factor model (χ^2 = 108.45, AIC = 154.45, CFI = .91, SRMR = .11). Finally, a four-factor model with the following structure was estimated: Factor 1 (effective IQ, observational skills, propaganda skills) versus Factor 2 (emotional stability, social relations, security) versus Factor 3 (motiva-

relation to security with respect to personnel selection even today.[3] More speciÞcally, for example, one can extend this concern to an assessment focus on forms of interpersonal dysgreater reliability and precision in terms of the EFA approach. The CFA approach used here, moreover, provided a powerful approach to the OSS data that enabled this study to home in on a model that provides the best fit to these unique data. The three-factor model presented here might be useful in other discussions in the intelligence community where personnel selec-