International Journal of Listening
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653656

Revisiting the Listening Styles Profile (LSP-16): A Confirmatory Factor Analytic Approach to Scale Validation and Reliability Estimation

Graham D. Bodie [a];Debra L. Worthington [b]
[a] Department of Communication Studies, The Louisiana State University, [b] Department of Communication and Journalism, Auburn University,

Online publication date: 06 May 2010

To cite this Article Bodie, Graham D. andWorthington, Debra L.(2010) 'Revisiting the Listening Styles Profile (LSP-16): A Confirmatory Factor Analytic Approach to Scale Validation and Reliability Estimation', International Journal of Listening, 24: 2, 69 — 88

To link to this Article: DOI: 10.1080/10904011003744516
URL: http://dx.doi.org/10.1080/10904011003744516

THE INTL. JOURNAL OF LISTENING, *24*: 69–88, 2010

ISSN: 1090-4018 print / 1932-586X online
DOI: 10.1080/10904011003744516

Revisiting the Listening Styles Profile (LSP-16): A Confirmatory Factor Analytic Approach to Scale Validation and Reliability Estimation

Graham D. Bodie
Department of Communication Studies
The Louisiana State University

Debra L. Worthington
Department of Communication and Journalism
Auburn University

The Listening Styles Profile (LSP-16) was developed to measure an individual's preferred listening style. One frequent criticism of the LSP-16 is the consistently low estimates of internal consistency. The following study addresses this concern using confirmatory factor analysis to assess both the latent constructs of the scale (i.e., People, Content, Action, Time) and the scale's reliability. Results suggest that listening style is multidimensional; however, additional scale development is needed to increase subscale reliability estimates. Suggestions for future research and development are provided.

Listening has been systematically studied for more than 50 years. During that time, a variety of listening concepts have been examined (e.g., listening fidelity, listening style, listening comprehension), and a variety of scales have been developed, with many of these scales measuring individual differences (for review, see Bodie & Fitch-Hauser, 2010). One particularly influential scale is the

---

A previous version of this manuscript was presented at the 2007 International Listening Association Convention.

Correspondence concerning this article should be addressed to Graham D. Bodie, Department of Communication Studies, The Louisiana State University, 136 Coates Hall, Baton Rouge, LA 70803. E-mail: gbodie@lsu.edu

Listening Styles Profile (LSP-16). Developed by Watson, Barker, and Weaver (1995), the LSP-16 was designed to measure an individual's primary preference for listening across four latent constructs (people-, action-, content-, and time-oriented listening). Although a number of studies have utilized the LSP-16 and found listening preferences related to a host of personality and other trait-like variables in line with theoretical predictions (Bodie & Villaume, 2003; Sargent, Fitch-Hauser, & Weaver, 1997; Villaume & Bodie, 2007; Weaver, 1998; Worthington, 2003, 2005), the reliability estimates of its subscales are consistently below acceptable values (Nunnally, 1978). While reliability is a product of sample and not an aspect of a particular scale (Thompson, 2003), the stable lack of reliability across samples has the unfortunate side-effect of potentially compromising the results of research utilizing the LSP-16 and thwarting efforts to advance listening theory (Bodie, 2009).

The goal of this article is to investigate potential reasons for low reliability estimates of the LSP-16. We begin with a brief review of the original conception of listening style as measured by the LSP-16 and assessments of its reliability and validity. Then, to provide a background for our data analytic strategy, we outline the steps of a confirmatory factor analysis prior to presenting our results. The discussion focuses on suggesting ways to improve the psychometric properties of the scale.

LISTENING STYLE PREFERENCE

Basing their research on previous studies that indicated most of us listen in habitual ways (Langer, 1980; Shiffrin & Schneider, 1977), Watson et al. (1995) reported that when listening to others, people seemed to have four primary listening style preferences — people, action, content, and time. Individuals who report a people-orientation toward listening tend to be other-focused, concentrating on the relational content of messages. Thus, people listeners are adept at cuing in on message content related to the emotions, feelings, and moods of others (Barker & Watson, 2000). This listening preference is associated with a number of related communication and psychological concepts, including a relationally oriented communication style (Bodie & Villaume, 2003), the Myers-Briggs *feeling* construct (Worthington, 2003), empathic tendencies (Weaver & Kirtley, 1995), and conversational sensitivity (Chesebro, 1999); other research (Villaume & Bodie, 2007; Weaver, 1998; Weaver, Watson, & Barker, 1996) reports extraversion positively associated with the people style. This style seems to describe "a socially adept personality that focuses on the other person in a communicatively competent fashion" (Villaume & Bodie, 2007, p. 117).

In contrast, action-oriented listeners focus attention on the inconsistencies and errors in a speaker's message, often listening in what Barker and Watson (2000) term "outline form." Consequently, these listeners prefer presentations that aid them in achieving this goal. Specifically, they tend to preference speakers that construct organized, direct, and logical presentations (Keyton & Rhodes, 1994).

Content listeners are said to differ from action listeners in several ways. For example, content listeners believe it is important to listen fully to a speaker's message prior to forming an opinion about it (while action listeners tend to become frustrated if the speaker is "wasting time"). Of the four listening styles, content listeners are likely the most comfortable with listening to complex, technical information (Keyton & Rhodes, 1994), and they generally do not have a problem asking a speaker to clarify or provide additional support for claims. Reflecting the positive association between an individual's need for cognition and this listening style (Worthington, 2008), content listeners pay attention to the details and supporting evidence of a message (Barker & Watson, 2000).

Finally, as the name implies, time-oriented listening suggests a focus on interactions between time and message reception. Behavioral manifestations of this orientation include "clock watching," letting speakers know about listening time constraints and parameters (Watson et al., 1995), and interrupting others when pressed for time (Watson & Barker, 1992). Previous researchers have described time-oriented listeners as individuals who engage in "hurried interactions" (Sargent et al., 1997) and as "communicative time managers" (Worthington, 2001).

LSP-16 Reliability and Validity

The foundation of identifying and measuring any construct involves both reliability and validity. As noted previously, the relationship between the four listening styles and a variety of other constructs has been explored via the LSP-16. This research is part of the continuing process of establishing the validity of individual conceptualizations of listening style. This research is, however, only as good as the reliability estimates of the scales for the sample under question. Indeed, the relationship between listening style and other constructs can be no higher than the reliability coefficient produced in a given sample (Crocker & Algina, 1986). Published research (Bodie & Villaume, 2003; Chesebro, 1999; Keyton & Rhodes, 1994; Villaume & Bodie, 2007; Watson et al., 1995; Weaver, 1998; Weaver & Kirtley, 1995; Weaver et al., 1996; Worthington, 2003, 2005) reports Cronbach alpha values of the people subscale between .48 and .75. In these same studies, reported reliability for the action subscale has ranged between .59 and .71; estimates of the time subscale have ranged between .38 and .67; and estimates of the content listening style have ranged between .46 and .71.

Nunnally (1978) is often cited as the premier source on "acceptable" reliability estimates. Although frequently quoted as suggesting that reliability estimates should consistently be above .70 for the scale to demonstrate adequate internal consistency (Lance, Butts, & Michels, 2006), this criterion only pertains to research “in the early stages” and applies only when a researcher wants to save “time and energy by working with instruments that have only modest reliability” (Nunnally, 1978, p. 245). Nunnally continues by suggesting that reliability estimates of .80 are potentially adequate if the central concern of the research is with correlations or differences in group means. If the research is to take on applied meaning (i.e., if we are to make suggestions to individuals based on their listening style), then reliabilities should be as high as .95. In fact, when dealing with important recommendations or when attempting to make strong theoretical claims, any amount of error will influence results (Carmines & Zeller, 1979). In other words, regardless of claims made by some (e.g., see Kirtley & Honeycutt, 1996, p. 178; Watson et al., 1995, p. 7), a reliability estimate in the mid-sixties is not “adequate.” Low estimates are especially troublesome given that training efforts are often tailored to a person’s listening-orientation scores (Barker & Watson, 2000). Thus, understanding the potential sources of low reliability for the LSP-16 is important.

One chief claim advanced by Watson et al. (1995) is that the LSP-16 factors into four primary listening orientations. These orientations are said to be orthogonal, and items are meant to load on only one factor. The original scale was constructed through the use of principle components analysis (PCA), which is an exploratory method used to guide theory development and is most appropriately used when the researcher has no *a priori* assumptions about the relationship between and among the data (i.e., correlations and covariations). Certainly, then, Watson et al. were warranted in their use of PCA since their article was concerned solely with developing a multidimensional scale to assess listening preference. The authors unfortunately stopped short of providing independent evidence of the validity of their factor structure. Consequently, subsequent research becomes suspect insofar as it has assumed a factor structure that has yet to be empirically tested. In the literature on scale development, exploratory methods like PCA should be, at minimum, followed by confirmation studies that test the hypothesis that the factor structure can be replicated with independent data (DeVellis, 2003; Thompson, 2004). In this article we report the first confirmatory factor analysis (CFA) of the LSP-16, providing a long overdue, deductive test of its factor structure.

The second goal of this article is to use CFA to aid in the interpretation of alpha. Like any test statistic, the use of Cronbach’s alpha is dependent on meeting several assumptions (Becker, 2000; Komaroff, 1997; Miller, 1995; Shevlin, Miles, Davies, & Salker, 2000). The first assumption is that the scale consists of only one dimension. That is, each subscale of the LSP-16 should be supported by the CFA allowing for the use of four separate measures of internal consistency. The second assumption deals with the nature of individual scale items. Specifically, (a) each item should reflect one latent construct, and (b) all items should be equally valid construct indicators. This second assumption rarely holds in practice since factor loadings often vary in magnitude (Thompson, 2004). Typically, then, alpha is said to be a lower-bound estimate of the reliability of a scale within a particular sample (Thompson, 2003). Moreover, if the assumptions of alpha are not met, then it is not the most appropriate measure of internal consistency (Shevlin et al., 2000), and other estimates of scale reliability should be reported.

Thus, the current study reports our attempt to confirm the underlying factor structure of the LSP-16. We also compare this four-factor structure with other, potential structures to assess whether the scale is, in fact, multidimensional (and if so, what is the most parsimonious data structure). Finally, we investigate reasons for low alpha estimates by testing the assumptions of this test statistic as it is used to assess the internal consistency of the LSP-16.

METHODS

Participants and Procedure

Raw data were obtained from the authors of several published and unpublished research reports (Worthington, 2001, 2003, 2005, 2008; Worthington & Fitch-Hauser, 2004) examining the relationships among a variety of listening, communication, and psychosocial constructs. Thus, our study utilizes existing data from college student participants (N = 710). Volunteers were enrolled in introductory communication courses at Auburn University and were rewarded with a modest amount of extra credit for their participation. The average age of volunteers in the combined data set was 21.9 (SD = 2.26), with slightly more men (57.3%), than women (42.7%) participating. All participants reviewed an informed consent statement and completed the original LSP-16 measure.

Instrument

The LSP-16 (Watson et al., 1995) is a self-administered, 16-item scale designed to measure people-, action-, content-, and time-oriented approaches to listening and receiving information. Based on a five-point scale ranging from (0) “never” to (4) “always,” respondents specify their perception of how well each statement applies to them. Individual responses to each item are then averaged, with higher scores indicating a stronger preference for a particular listening style (see Table 1 for individual items).

Data Analytic Strategy

This study utilizes confirmatory factor analysis in order to test several assumed properties of the LSP-16. The process of CFA involves several steps: (1) model fit, (2) model comparison, (3) model respecification, and (4) parameter estimation.

TABLE 1
LSP-16 Items and Estimates of Factor Loadings and Variances for LSP-16 Model

Construct	*Item*	*Variance*	*Standard Error of Variance*	*Standardized Factor Loading*
ALS		.17	.04	—
	1. I am frustrated when others don't present their ideas in an orderly, efficient way.	.69	.04	.40
	2. When listening to others I focus on any inconsistencies and/or errors in what's being said.	.66	.04	.45
	3. I often jump ahead and/or finish thoughts of speakers.	.57	.04	.42
	4. I am impatient with people who ramble on during conversations.	.51	.05	.65
CLS		.58	.11	—
	5. I prefer to listen to technical information.	.65	.06	.53
	6. I prefer to hear facts and evidence so I can personally evaluate them.	.55	.03	.29
	7. I like the challenge of listening to complex information.	.35	.10	.79
	8. I ask questions to probe for additional information.	.54	.03	.24
PLS		.17	.03	—
	9. I focus my attention on the other person's feelings when listening to them.	.32	.03	.67
	10. When listening to others I quickly notice if they are pleased or disappointed.	.37	.03	.57
	11. I become involved when listening to the problems of others.	.42	.03	.52
	12. I nod my head and/or use eye contact to show interest in what others are saying.	.37	.02	.44
TLS		.37	.04	—
	13. When hurried I let the others know that I have a limited amount of time to listen.	.44	.03	.65
	14. I begin a discussion by telling others how long I have to meet.	.46	.03	.52
	15. I interrupt others when I feel time pressure.	.34	.03	.72
	16. I look at my watch or clocks in the room when I have limited time to listen to others.	.65	.04	.52

ALS = Action-oriented Listening; CLS = Content-oriented Listening; PLS = People-oriented Listening; TLS = Time-oriented Listening. All variance estimates were significant at $p < .001$. Variance estimates associated with items are error variances associated with that item, not item variances. All unstandardized factor loading estimates were significant at $p < .001$; exact unstandardized values are available from the first author upon request.

Model Fit

First, a model is specified to which the data are assumed to conform. Basically, a proposed model consists of independent and dependent variables. The dependent variables (those being predicted or influenced) are the individual scale items. Each scale item is influenced by two independent variables, namely a latent construct and an error term. In the present case, the latent constructs are the four listening styles, whereas error refers to the amount of variance in an item not explained by the latent construct. Ideally, each item should be influenced to a greater extent by the latent construct than by its error term; if so, evidence for construct validity is obtained — the LSP-16 items are capturing their supposed factors and are not measuring something they are not supposed to measure. Once the number of factors, the items proposed to load on their respective factors, and the interrelationships among the factors are specified, the computer program assesses how well that proposed structure is captured by the data.

For the purposes of this study, "goodness-of-fit" is tested using a variety of test statistics (Hoyle, 2000; Kline, 2005). Because of its standardized distribution and use in model comparison, and because all other fit indices are derived from this statistic, chi-square is reported. Contrary to null hypothesis significance testing, a significant chi-square indicates the model does a poor job of replicating the data covariance.[1] Chi-square is, however, sensitive to sample size. Thus, a chi-square to degrees of freedom fit ratio, which norms the value of the chi-square to sample size, is also reported. Values as low as 2.0 and as high as 5.0 have been recommended, indicating reasonable fit (Bollen, 1989). In addition, the comparative fit index (CFI) is reported to assess the relative improvement of the LSP-16 model as compared with the independence model which assumes zero population covariances among the items. Values above .90 indicate relatively good fit as compared to the null model (Hu & Bentler, 1999). The final index reported is the Root Mean Square Error of Approximation (RMSEA), which estimates the amount of error of approximation per model degree of freedom taking into account sample size. Values below .05 — and precise confidence intervals that include low values — are generally desirable (Browne & Cudeck, 1993).

In addition to single number estimates, we also inspected the residual covariance matrix, which captures the discrepancy between the theoretical model and the

[1]This makes sense when considering the purpose of the chi-square statistic: to compare an observed matrix to an expected matrix. A nonsignificant chi-square, therefore, suggests that the observed matrix is the same as the expected (i.e., theoretical) matrix. In other words, the sample data conform to the proposed factor structure.

covariance structure of the data.[2] Overall, the residual covariance matrix can be thought of as a measure of model fit for variances and covariances as opposed to a single number fit statistic (Raykov & Marcoulides, 2006). In other words, a model can be specified correctly in some ways but misspecified in other ways; the residual matrix provides information of model misspecification.[3] Thus, the values provided in the standardized residual matrix represent the inconsistency between the proposed model (the four-factor LSP-16 model) and what the data suggest about how the individual items actually covary among each other in our sample.[4] As the value for any standardized residual covariance increases, there is an increasingly poor fit for the two items in question. The more large residuals there are, the less well the data fit the theoretical model. Raykov and Marcoulides (2006) recommend residual values of two or greater "[indicate] that the model considerably underexplains a particular relationship between two variables," while values of negative two or smaller "generally [indicate] that the model markedly overexplains the relationships between the two variables" (pp. 48–49).

Model Comparison and Respecification

CFA can be used to test competing or alternative models. Thus, in our example, listening styles could be represented by a single factor or by two or three factors (as opposed to four). In fact, previous research strongly suggests that the people-oriented listening style is the only "pure" style, with the other styles combining to predict other personality-like and communication predispositions (Bodie & Villaume, 2003; Villaume & Bodie, 2007). Thus, CFA allows us to answer whether (a) the four-factor model represents the data in the most parsimonious way or (b) correlations among the four factors are explained by a higher-order factor called "listening preference." This latter model (a second-order model) is, in fact, what is implied by Watson et al. in their original conceptualization of preferences for receiving information. Depending on which model reproduces

[2]For those familiar with exploratory factor analytic methods, the residual matrix can be reproduced within those programs as well. In fact, within the framework of PCA, the model that converges has the lowest possible values in the residual matrix (which is why some factor analyses take more iterations to converge).

[3]In the language of regression, a residual is, for any given data point, the difference between the value of the dependent variable and the predicted value of the dependent variable as estimated from one or more independent variables. As in regression where the dependent variable is never perfectly predicted by the set of independent variables, in CFA the theoretical model is rarely perfectly predicted by the sample data.

[4]Basically, if all values equal zero then the sample data exactly replicated the theoretical model and chi-square would also equal zero. When chi-square is not zero (and is significantly different from zero), we can expect the residual covariances to deviate from zero as well.

the data covariance structure with the least residual error, the LSP-16 may have to be respecified to conduct estimates of parameter values (see section below).

Parameter Estimation

Overall model fit provides evidence of construct validity. In addition, the individual items should be valid indicators of their assumed factor. Within PCA, researchers typically use rules such as the Fürntratt criteria to interpret item loadings (e.g., Imhof & Janusik, 2006). Although similar criteria exist for CFA, this approach restricts items to load on only one factor; if cross loadings exist, they will be represented as error in the model.

When items are purported to load on only one factor, high standardized factor loadings for all items (> .60) (Hair, Anderson, Tatham, & Black, 1998) provide an indication of convergent validity; in other words, all items load appropriately on this factor. Depending on how well certain items "load" on their latent construct and whether all items load equally well, estimates of alpha are interpreted differently. As we specified above, the use of Cronbach's alpha depends on meeting several assumptions. We test these assumptions in the framework of CFA by inspecting parameter estimates and the overall fit of models that allow parameter estimates to be equivalent versus the overall fit of models that allow parameter estimates to freely vary.

Finally, CFA estimates the error associated with each item and allows inspection of covariation among error terms of individual items. If the error terms of individual items are related in a systematic manner, this can indicate that items are measuring more than one underlying concept or that similar systematic sources of variability are influencing these items (Gerbing & Anderson, 1984). Alternatively, if error terms are high but unrelated, there is a good deal of random (measurement) error in the model.

RESULTS

Assumptions underlying multivariate methods were inspected (Tabachnick & Fidell, 2007), and several data points were found problematic.[5] The final data set consisted of 661 individuals with no missing data. Based on recommendations by Hu, Bentler, and Kano (1992), the study was sufficiently powered to assess model fit and provide parameter estimates. Corresponding with past research, low internal consistency as assessed by Cronbach's coefficient alpha was found for the four listening styles (People, α = .65; Action, α = .53; Content, α = .52; Time, α = .67).

[5]All analyses are available from the first author upon request.

Model Fit for LSP-16

The theoretical model proposed by Watson et al. (1995) and assumed in subsequent research was tested using maximum likelihood estimation in Amos 16.0 (SPSS, 2007).[6] The original four-factor model of People, Content, Action, and Time contains four items for each factor and allows the styles to be correlated (estimates of factor correlations are presented in Table 2). As seen in Table 3, although the fit ratio and RMSEA were within range of an adequate fitting model, CFI was below .90 suggesting that the LSP-16 is accurately specified in some areas but misspecified in others.

To examine where the model fails to capture the covariance structure of the sample data, the residual covariance matrix was examined. For this model, there

TABLE 2
Estimated Correlations for the LSP-16

	1	*2*	*3*
1. People	—		
2. Action	−.07	—	
3. Content	.06	.10	—
4. Time	−.01	.45*	−.05

*p < .001.

TABLE 3
Fit Index Estimates for the LSP-16 and Alternative Models

χ^2	*df*	*p*	*Fit Ratio*	*CFI*	*RMSEA*	*RMSEA 90% CI*	
Four-Factor Model							
312.96	98	< .001	3.19	.84	.06	.05	.07
One-Factor Model							
891	104	< .001	8.57	.41	.11	.10	.11
Two-Factor Model							
542.52	101	< .001	5.37	.67	.08	.08	.09
Three-Factor Model							
423.44	99	< .001	4.28	.76	.07	.06	.08
Second-Order Model							
317.76	100	< .001	3.18	.84	.06	.05	.07
Tau-Equivalent Model							
417.83	110	< .001	3.80	.77	.07	.06	.07

[6]Maximum likelihood method uses a fitting function analogous to the least squares criterion in regression.

are 120 possible residual covariances, each one representing the misfit of the data to the proposed model for two of the scale items. For our data, 24 of the standardized residual covariance values (20%) were above 2 in absolute value. Of these 24 values, 18 were positive and 6 were negative. Thus, the LSP-16 seems to underexplain the relationship among 18 sets of items and overexplain the relationship among 6 sets of items. In each case, the primary cause of the misspecification was the content-oriented listening style. Specifically, the items "I prefer to hear facts and evidence so I can personally evaluate them." and "I ask questions to probe for additional information." were the cause of 12 of the 18 values (67%) over positive 2, whereas the item" I prefer to listen to technical information." was responsible for four of the six values (67%) greater than negative 2. As will be shown in a subsequent section, these three content-oriented listening style items also did not load highly on their predicted latent factor and had high error variances.[7]

Model Comparison

For any given data set, an infinite number of possible models exist that are empirically indistinguishable (Cliff, 1983; MacCallum, Wegener, Uchino, & Fabrigar, 1993). Fortunately, only a subset of these models is theoretically warranted, limiting the need to run thousands upon thousands of models. For the current data, we conducted a series of model comparisons to ascertain whether listening preference is more parsimoniously represented by another factor structure that also makes theoretical sense. Before getting to these models, it seemed appropriate to test the most parsimonious model for this data, a one-factor model. Supporting the contention that listening style is a multidimensional construct, this model produced overall poor model fit (see Table 3) and was appreciably worse than any of the other alternative models, all *p*s < .001.[8]

Given that previous research suggests people-oriented listening is the only "pure" style, a two-factor model was specified by constraining the correlations between the latent factors representing action-, time-, and content-orientations and leaving the people-orientation to freely vary. As seen in Table 3, this model produced an appreciably worse model fit as indicated by a large increase in chi-square, $\Delta\chi^2$ (3) = 229.56, $p < .001$, and decrease in CFI, ΔCFI = .17.

[7]Indeed, removal of the content factor, all four content items, and the item error terms produced a well-fitting model, χ^2 (51) = 124.91, $p < .001$, fit ratio = 2.45, CFI = .93, RMSEA = .05 (CI90% = .04, .06).

[8]For all alternative (and nested) models, the way in which we determined one model as "better" than another was by comparing the two chi-square values. Two models can be compared by looking at a chi-square change score that is generated by subtracting the smaller of the two chi-square values from the larger (Kline, 2005). The statistical significance of this value is determined by comparing the change value to a critical value found in a chi-square table (Tabachnick & Fidell, 2007).

Other research suggests that correlations among action-, content-, and time-oriented listening may be a function of these orientations combining in one or more ways in order to produce three, as opposed to four, listening styles. In the current sample, however, the only two factors significantly correlated were the time- and action-oriented styles (see Table 1). Thus, a three-factor model was tested by constraining the correlation between action- and time-oriented listening while allowing the people- and content-oriented styles to freely vary. The three-factor model was appreciably better than the two-factor model, $\Delta\chi^2$ (2) = 119.08, $p < .001$; however, it was not an improvement over the original four-factor model, $\Delta\chi^2$ (1) = 110.48, $p < .001$.

The final model tested was a second-order model that explains the correlations among the four listening styles as the result of a superordinate factor called *listening style*. This model produced a chi-square statistic indistinguishable from the original LSP-16 model, $\Delta\chi^2$ (1) = 4.65, $p > .05$; thus, it should be preferred over the original model since it is simpler. Unfortunately, inspection of other fit indices as well as parameter estimates associated with the time-oriented listening style caused us to reject this model. First, none of the second-order loading estimates (i.e., how well each of the four styles loaded on the superordinate listening style factor) achieved statistical significance. Second, the error variance associated with the time-oriented style was negative. Finally, many of the standardized residual covariances increased in value, with several achieving absolute values over five. Ultimately, therefore, a model specifying four correlated latent factors is preferred.[9] Given this, we turn our attention to examining the individual items and modification indices of the original LSP-16 model in an effort to explain (a) why the current data do not match the originally proposed four-factor model and (b) why internal consistency estimates of the LSP-16 are poor.

Parameter Estimates

The last column of Table 1 shows the standardized factor loadings for all 16 items. These values are interpreted in a manner similar to standardized factor loadings in a standard, unrotated PCA: squaring standardized estimates provides an estimate of the latent construct variance accounted for by an item which can be interpreted as the reliability of a given item analogous to the communality coefficient (h^2) (Shevlin et al., 2000). As seen in Table 1, standardized regression

[9]As one reviewer suggested, although the "four-factor model most accurately captures listening styles among the models examined [it] could be that a model with more than four factors would be a better fit for the data (even though there was no theoretical reason to test for such a model in this case)." To test this assertion, we used several data driven methods. None of the exploratory factor analyses suggested the extraction of more than four factors. In fact, all exploratory analyses suggested the interpretation of four factors. All analyses are available from the first author upon request.

weights for most of the items are below .71 (the equivalent of a .50 reliability coefficient for that item). In fact, only 6 of the factor loadings are considered “high” in magnitude ($\lambda > .60$) (Hair et al., 1998).

With CFA models, each item is not only specified to load on a latent construct, but each item also has an error component to which it is associated. The variance of these error components is also estimated, providing an estimate of the variance in the item not explained by the latent variable (i.e., measurement error or unique variance to that item). As seen in Table 1, 4 of the 16 error variances are above .60, four are between .51 and .60, and an additional 3 are between .40 and .50. These high error variances suggest that items are either unreliable indicators of particular factors or there are some common, systematic sources of variability shared among many of the items. Since the second-order analysis presented above in the model comparison section failed to reduce these error variances, it is likely that this error is random. Inspection of the correlations among error terms presented in Table 4 suggests the plausibility of this conclusion since only nine of the 120 correlations were statistically significant (at $p < .05$) and only one was even moderate in magnitude ($r > .30$) (Cohen, 1988). Moreover, seven of the 16 error variances are greater than the standardized loadings. Further evidence of random error comes from the fact that the average correlations between individual items are all below .40 (action, $r_{ave} = .23$; content, $r_{ave} = .20$; people, $r_{ave} = .30$; time, $r_{ave} = .36$; see Table 5).

Internal Consistency

As stated above, the typical measure of internal consistency reported for the LSP-16 is Cronbach’s coefficient alpha. This test statistic, however, has been used without appropriate tests of the assumptions underlying it. Specifically, alpha assumes a measurement model that is tau-equivalent. Tau-equivalence means that each item is an equivalent embodiment of its latent construct. To test this assumption, we constrained the factor loading of each item to one, in turn telling the program to assume those values are equal (Hoyle, 2000). The fit of this model was appreciably worse than the original model, $\Delta\chi^2$ (12) = 104.87, $p < .001$ (see Table 2 for all fit statistics). Thus, the LSP-16 is, at best, congeneric (i.e., items are measuring a construct with different levels of precision and error), although low factor loadings put this assumption in question. With congeneric models, alpha is not the most accurate measure of internal consistency reliability; for these models, alpha tends to underestimate the actual reliability of a scale within that sample. Instead, Raykov (2001, 2004) suggests using an alternative point estimate for scale reliability. This measure of reliability, ρ, is analogous to Cronbach’s alpha insofar as it measures the internal consistency of a set of items; however, Raykov’s measure computes this estimate using the factor loadings and

TABLE 4
Correlations of Error Terms in the LSP-16 Model

	e1	*e2*	*e3*	*e4*	*e5*	*e6*	*e7*	*e8*	*e9*	*e10*	*e11*	*e12*	*e13*	*e14*	*e15*
e1	1														
e2	−.003	1													
e3	−.003	−.004	1												
e4	−.124**	−.161**	−.132**	1											
e5		.018		−.002	1										
e6		.003			.045	1									
e7		−.042		.005	−.569**	−.106*	1								
e8		.002			.033	.006	−.077	1							
e9								−.027	1						
e10								−.01	−.185**	1					
e11								−.007	−.132**	−.05	1				
e12								−.004	−.081	−.03	−.022	1			
e13								−.003					1		
e14								−.001					−.033	1	
e15					.001		−.001	−.009					−.233**	−.088*	1
e16								−.001					−.033	−.012	−.087

The number beside the “e” represents the item listed in Table 1. So, for example, e1 is the error for the item “I am frustrated when others don’t present their ideas in an orderly, efficient way.” Blank spaces within the table represent correlations that were zero to the third decimal place. *$p < .05$, **$p < .01$.

TABLE 5
Correlations Among Scale Items

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
1	1														
2	.21**	1													
3	.10*	.22**	1												
4	.30**	.26**	.27**	1											
5	.01	.08	–.03	–.01	1										
6	.04	.14**	.02	.09*	.18**	1									
7	–.01	.13**	.06	–.02	.43**	.20**	1								
8	.07	.13**	.09*	.03	.03	.17**	.21**	1							
9	.03	–.13**	–.07	–.08	–.12**	.07	–.01	.14**	1						
10	.04	.05	.01	.04	–.04	.15**	.09*	.20**	.39**	1					
11	.02	–.01	.08	–.07	–.10*	.05	.04	.15**	.34**	.28**	1				
12	.10	–.01	.04	–.02	–.03	.16**	.01	.18**	.28**	.23**	.28**	1			
13	.01	.09*	.11*	.19**	–.03	.05	.02	.18**	–.01	.10*	.00	.03	1		
14	.11**	.03	.06	.09*	–.01	–.07	.03	.05	.01	–.02	.03	–.02	.37**	1	
15	.13**	.13**	.20**	.22**	–.12	–.03	–.07	.12**	–.07	.02	–.01	.06	.46**	.38**	1
16	.09*	.17**	.20**	.24**	–.05	.02	–.06	.05	–.06	.03	.02	-.01	.33**	.26**	.35**

Numbers correspond to the items as listed in Table 3. $*p < .05$, $**p < .01$.

error terms generated through CFA and does not make assumptions about tau-equivalence.[10] For each LSP-16 subscale, the reliability coefficients using Raykov's measure were, as expected, slightly higher than the estimates generated using alpha (ρ_{ALS} = .62, ρ_{CLS} = .57, ρ_{PLS} = .64, ρ_{TLS} = .69); they were still, however, not acceptable using the standard set by Nunnally (1978).

DISCUSSION

This study sought to provide the first confirmatory factor analysis of the LSP-16 to test the validity of its factor structure originally produced through exploratory methods and to provide a more informed analysis of internal consistency reliability estimates reported in the literature using this scale. CFA allowed us to test the proposed LSP-16 model and how it compared to empirically equivalent models. The analyses reported above also provided insight into why the reliability estimates for the subscales are consistently below acceptable values. Based on an inspection of several fit statistics, we can conclude that the original LSP-16 model fit the covariance structure of our sample data better than alternative models. Unfortunately, this multidimensional scale has serious psychometric problems. Inspection of the residual covariance matrix indicated that while six item-item relationships were overestimated, the LSP-16 model primarily underestimates relationships between items. In other words, and in line with prior research (Bodie & Villaume, 2003; Villaume & Bodie, 2007), the four styles as represented by the LSP-16 are not "pure" but are related in ways that have yet to be fully appreciated in most explanations of listening styles to date.

Although model misspecification was primarily localized to the content style, the presence of high error variances for all items and the low level of covariation among error variances indicate that there is a substantial amount of measurement error present in the LSP-16. This measurement error, along with the misspecification of the tau-equivalent model, is the primary contributor to consistently low estimates of Cronbach's coefficient alpha. In addition, the alternative measure of

[10]The formula used for the point estimation of scale reliability was

$$\rho_Y = \frac{(\Sigma\lambda_i)^2}{[\,(\Sigma\lambda_i)^2 + \Sigma\theta_{ii}\,],}$$

where $(\Sigma\lambda_i)^2$ is the squared sum of unstandardized regression weights and $\Sigma\theta_{ii}$ is the sum of unstandardized measurement error variances.

internal consistency still suggests that the subscales of the LSP-16 are not reliable measures of any of the four primary listening styles.

Overall, several conclusions and recommendations can be made from the data presented above. First, this study is consistent with past empirical work that has consistently reported low reliability estimates of the LSP-16. Although reliability is a product of individual studies and not a property of a scale, the confirmatory approach taken calls into question claims made about the scale's construct validity. Certainly, listening style appears to be a multidimensional construct, as supported by the four-factor model producing a better fit than a one-factor model (see Table 3). The four-factor model was also a better fit to the sample data than the alterative models tested, further supporting the claim that the four-factor model most accurately captures listening styles as measured by the LSP-16. At the same time, however, other CFA results suggest that the scale does not fully operate as intended. At present, we recommend rewriting items and testing a revised version of the scale as the next step needed to enhance the reliability of the LSP-16 and to further explore its construct validity. Examples of research with a focus on scale redevelopment can be found in Shearman and Levine (2006) and Levine et al. (2004).

Developing additional items designed to measure multiple dimensions of listening style is warranted for at least two reasons. First, improving individual scale reliability can be achieved by including more similarly worded items (DeVellis, 2003). Second, most of the factor loadings for the individual items indicated that items are not sharing much variability with their supposed latent construct. Writing items that do a better job of loading on latent constructs will, in turn, increase estimates of internal consistency. If future studies show that all items load on four similar latent constructs as the scale was originally constructed, then Watson et al.'s theoretical framework holds. If, however, newly developed items begin to load on separate factors (factors other than the four primary styles discussed to date), then the multidimensional nature of listening style may be valid, but the exact number of listening style factors will change.

One way in which to go about developing new items for an updated version of the LSP-16 is to inspect the nature of the original items, that is, their wording, how well they load on a latent construct, and the magnitude of their error variances. With respect to wording, it seems that some of the LSP-16 items are tapping a cognitive aspect of a particular style, whereas other items are tapping a behavioral one. For example, within the action-oriented listening style, items one and four seem to capture cognitions (what the person prefers or does that may or may not be noticeable to the speaker), whereas item three is clearly behavioral (see Table 1). On a conceptual level, this is not problematic, as listening is both cognitive and behavioral (Janusik, 2007). This distinction is, however, potentially problematic at the statistical level insofar as these dimensions (cognition and behavior) may represent separate components of each listening style. If this is an accurate depiction of what is occurring when people are filling out the LSP-16, future research should find that cognitively based items for each style are loading together and that behaviorally based items for each style are loading together. Statistically, revisions of the LSP-16 may benefit from implementing item parceling techniques (Matsunaga, 2008) that take into account the possibility of item clusters within individual listening styles. Thus, we recommend future item development attempting to generate an equal number of cognitively and behaviorally based items so that this hypothesis can be tested.

In addition to developing new items, another potential change to the scale might include changes in the scale's response options. For instance, changing response options from "never – always" to "strongly disagree – strongly agree" should provide further evidence of potential confounds of alpha reliability estimates. Perhaps, as in most aspects of life, people rarely operate on the extremes. The options "never" and "always" may be contributing to range restriction that can undermine variability among items and cause the scale to break down. Indeed, inspection of the frequencies of individual items for the present data shows that the "never" response was chosen infrequently for most items with the exception of one time item ("I begin a discussion by telling others how long I have to meet," 16%) and one content item ("I prefer to listen to technical information," 19.1%). Indeed, for several of the items, "never" was not chosen at all by any of the 661 respondents.

CONCLUSION

To ensure quality research and successful training, it is necessary for us to develop a measure of listening style that has a stable factor structure, that has been replicated, and that has been shown to be invariant across both populations (e.g., cultures) and applications. This article presents just the first step of a program of research that should be undertaken with respect to the LSP-16. We hope this initial study prompts much needed validation research not only for the LSP-16 but also for other scales designed to measure multidimensional listening constructs.

REFERENCES

Barker, L. L., & Watson, K. W. (2000). *Listen up*. New York: St. Martin's Press.

Becker, G. (2000). Coefficient alpha: Some terminological ambiguities and related misconceptions. *Psychological Reports, 86*, 365–372.

Bodie, G. D. (2009). Evaluating listening theory: Development and illustration of five criteria. *International Journal of Listening, 23*, 81–103.

Bodie, G. D., & Fitch-Hauser, M. (2010). Quantitative research in listening: Explication and overview. In A. D. Wolvin (Ed.), *Listening and human communication in the 21st century* (pp. 46–90). Oxford: Blackwell.

Bodie, G. D., & Villaume, W. A. (2003). Aspects of receiving information: The relationship between listening preferences, communication apprehension, receiver apprehension, and communicator style. *International Journal of Listening, 17*, 47–67.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, NJ: Sage.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage.

Chesebro, J. L. (1999). The relationship between listening styles and conversational sensitivity. *Communication Research Reports, 16*, 233–238.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18*, 115–126.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Holt, Rinehart and Winston.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research, 11*, 572–580.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis with readings* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 465–497). San Diego, CA: Academic Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*, 351–362.

Imhof, M., & Janusik, L. (2006). Development and validation of the Imhof-Janusik listening concepts inventory to measure listening conceptualization differences between cultures. *Journal of Intercultural Communication Research, 35*, 79–98.

Janusik, L. (2007). Building listening theory: The validation of the conversational listening span. *Communication Studies, 58*, 139–156.

Keyton, J., & Rhodes, S. C. (1994). The effects of listener preference styles on identifying sexual harassment. *Journal of the International Listening Association, 8*, 50–79.

Kirtley, M. D., & Honeycutt, J. M. (1996). Listening styles and their correspondence with second guessing. *Communication Research Reports, 13*, 174–182.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.

Komaroff, E. (1997). Effect of simultaneous violations of essential tau-equivalence and uncorrelated error on coefficient alpha. *Applied Psychological Measurement, 21*, 337–348.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*, 202–220.

Langer, E. (1980). Rethinking the role of thought in social interaction. In H. Harvey, W. Ickes, & R. Kidd (Eds.), *New directions in attribution research* (pp. 35–58). Hillsdale, NJ: Erlbaum.

Levine, T. R., Beatty, M. J., Limon, S., Hamilton, M. A., Buck, R., & Chory-Asada, R. M. (2004). The dimensionality of the verbal aggressiveness scale. *Communication Monographs, 71*, 245–268.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185–199.

Matsunaga, M. (2008). Item parceling in structural equation modeling: A primer. *Communication Methods and Measures, 2*, 260–293.

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling, 2*, 255–273.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology, 54*, 315–323.

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy, 35*, 299–331.

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbuam.

Sargent, S. L., Fitch-Hauser, M., & Weaver, J. B., III. (1997). A listening styles profile of the Type-A personality. *International Journal of Listening, 11*, 1–14.

Shearman, S. M., & Levine, T. R. (2006). Dogmatism updated: A scale revision and validation. *Communication Quarterly, 54*, 275–292.

Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Salker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences, 28*, 229–237.

Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing, II: Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.

SPSS. (2007). AMOS for Windows. Analysis of moment structures (Version 16.0). Chicago: SPSS.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education.

Thompson, B. (2003). Understanding reliability and Coefficient alpha, really. In B. Thompson (Ed.), *Score reliability: Contemporary thinking on reliability issues* (pp. 3–23). Thousand Oaks, CA: Sage.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: APA.

Villaume, W. A., & Bodie, G. D. (2007). Discovering the listener within us: The impact of trait-like personality variables and communicator styles on preferences for listening style. *International Journal of Listening, 21*, 102–123.

Watson, K. W., & Barker, L. L. (1992). Comparison of the ETS national teacher examination listening model with models used in two standardized tests. *International Journal of Listening, 6*, 32–44.

Watson, K. W., Barker, L. L., & Weaver, J. B. (1995). The listening styles profile (LSP-16): Development and validation of an instrument to assess four listening styles. *International Journal of Listening, 9*, 1–13.

Weaver, J. B., III. (1998). Personality and self-perceptions about communication. In J. C. McCroskey, J. A. Daly, M. Martin, & M. J. Beatty (Eds.), *Communication and personality* (pp. 95–117). Cresskill, NJ: Hampton Press.

Weaver, J. B., III, & Kirtley, M. D. (1995). Listening styles and empathy. *Southern Journal of Speech Communication, 60*, 131–140.

Weaver, J. B., III, Watson, K. W., & Barker, L. L. (1996). Individual differences in listening style: Do you hear what I hear? *Personality and Individual Differences, 20*, 381–387.

Worthington, D. L. (2001). Exploring juror listening processes: The effect of listening style preference on juror decision making. *International Journal of Listening, 15*, 20–35.

Worthington, D. L. (2003). Exploring the relationship between listening style preference and personality. *International Journal of Listening, 17*, 68–87.

Worthington, D. L. (2005). Exploring the relationship between listening style preference and verbal aggressiveness. *International Journal of Listening, 19*, 3–11.

Worthington, D. L. (2008). Exploring the relationship between listening style and need for cognition. *International Journal of Listening, 22*, 46–58.

Worthington, D. L., & Fitch-Hauser, M. (2004, November). *Listening and learning: Exploring the relationship between student listening style preference and learning style*. Paper presented at the annual meeting of the National Communication Association, Chicago, IL.