

Articles

Perceptual Learning Styles and Lessons in Psychometric Weakness

Ian Isemonger
Kumamoto University

In this critical review, I argue that the usefulness of perceptual learning styles constructs within applied linguistics is very limited. Researchers within applied linguistics have neglected to engage with objections to these constructs which date back to the 1970s within general educational research. Problems of poor instrumentation are considerable and predictive power has not sufficiently been demonstrated. It is argued that these constructs present a special case for measurement because they are not easily operationalized through statements on self-report questionnaires. I discuss implications for practitioners and research, and argue for greater editorial oversight in preventing poor instruments from entering the literature in the future. Some specific recommendations which may assist with such prevention are discussed. These include a more critical approach to the use of Cronbach's alpha, the use of Confirmatory Factor Analysis (CFA) as one powerful tool to demonstrate unidimensionality, and the avoidance of paraphrased items.

本論文は、応用言語学で用いられる知覚学習スタイルの構成概念が有用性に欠けることを指摘する。一般的な教育学研究の分野で1970年代から批判されてきたこれら構成概念について、応用言語学研究者たちはその対応法の検討を十分には行なっていない。知覚学習スタイルの測定方法には問題が多く、その妥当性も明確になっていない。さらに、アンケートのような自己報告を用いた測定方法では、このような構成概念は適切に定義づけることが難しいことも指摘されている。外国語教育実践と応用言語学研究のためにも、不十分な測定方法が将来的に研究に入り込まないようにする必要があり、そのためには学術誌等の編集者によるさらなるチェック体制の強化が不可欠である。具体的方法として、クロンバック α 係数の使用に関する注意喚起や、1次元性を確認するのに有用な確証的因子分析(CFA)の活用を提案する。

Perceptual learning styles emerged as a significant branch of study within applied linguistics during the 1980s. This was led by a groundbreaking paper (Reid, 1987) using the Perceptual Learning Styles Preference Questionnaire (PLSPQ). There was a surge of interest in the area with a number of studies employing the PLSPQ (Frank & Hughes, 2002; Hyland, 1993; Isemonger & Sheppard, 2003; Kelly, 1998; Kinsella, 1995b; Melton, 1990; O'Donoghue, Oyabu, & Akiyoshi, 2001; Oxford, 1995; Oxford & Anderson, 1995; Peacock, 2001; Shen, 2010; Siew Luan & Ngoh, 2006; Stebbins, 1995; Thomas, Cox, & Kojima, 2000; Yamashita, 1995; Yu-rong, 2007). This interest was arguably driven by the intuitive or common-sense appeal of perceptual learning styles as a potential area of individual and cross-cultural differences. Also, shortly after Reid's release of the PLSPQ, three further perceptual learning styles instruments emerged. The first of these, the Learning Channel Preference Checklist (LCPC; O'Brien, 1990) was revised by the same author 12 years later (O'Brien, 2002). The second, the Style Analysis Survey (SAS; Oxford 1993a; 1993b) measures other aspects of learning styles in addition to preferences for perceptual learning styles. The third, the Perceptual Learning Preference Survey (PLPS; Kinsella, 1995a) has far less exposure in the literature. In addition to these earlier instruments, DeCapua and Wintergerst (2005) and Wintergerst, DeCapua, and Verna (2002; 2003) have more recently sought to revise the PLSPQ in the form of the Learning Styles Indicator (LSI).

Unfortunately, the initial appeal of perceptual learning styles research has not been matched by tangible gains for applied linguistics. This disappointment is the result of a general omission to engage with the predictive power of these constructs in sound empirical studies and, above all, to address the psychometrics of scores generated by scales purporting to measure the constructs themselves—an issue which methodologically precedes the issue of predictive power. Twenty-five years on from the publication of Reid's (1987) paper, my purpose in this critical review is to scrutinize the emergence and development of the ensuing research trajectory and its associated line of instrumentation. A case is presented for where it went wrong and what can be learned from some of the mistakes. Such an assessment is important for future areas of research growth within applied linguistics and is also important in alerting both researchers and practitioners to the apparent limitations of these constructs.

The paper begins by drawing the reader's attention to the remarkable manner in which the weaknesses of perceptual learning styles research within applied linguistics parallel weaknesses and controversy within gen-

eral educational research occurring about a decade earlier. This brief detour into an instructive past, neglected by perceptual preference researchers within applied linguistics up to now,¹ is offered to strengthen the arguments for a change of thinking on the usefulness of these constructs for both research and classroom practice. Following this, the emergence of perceptual learning styles instruments within applied linguistics is considered from the perspective of psychometric credibility and predictive power. In a more positive mode of critique, suggestions for editorial oversight that might help prevent such weakness in the future are offered, and the paper then turns to some specific issues and associated guidelines for promoting more empirically secure instrumentation and its inclusion in the literature in the future. These suggestions, while important, are not intended to be exhaustive or a comprehensive treatment of the issues. Finally, attention is drawn to the above-mentioned attempts by DeCapua and Wintergerst (2005) and Wintergerst et al. (2002; 2003) to revise the PLSPQ in the form of the LSI. It is argued that their approach risks compounding the problems by retaining some of the inherent flaws of the PLSPQ and thereby perpetuating the life of an instrument, and indeed a line of instrumentation, that should be consigned to applied linguistics' psychometric past.

Preferences for Perceptual Modalities and Repeating History

The initial interest in perceptual learning styles within applied linguistics almost 25 years ago was at least partly informed by theory and constructs already in use concerning preferences for different modalities of perception. These constructs included the visual, auditory, and tactile modalities of perception and were initially operationalized through self-report by R. Dunn and K. Dunn in the 1970s in work that has covered three or four decades. This work is represented in the following contemporary instruments: the Learning Styles Inventory (Price & Dunn, 1997) or LSI (not to be confused with the Learning Styles Indicator cited above) and the Productivity Environmental Preference Survey (PEPS; Price, Dunn, & Dunn, 1996). The LSI and PEPS are designed to measure more than perceptual constructs, but the perceptual constructs contained in them prompted Reid's research (Dunn, 1983, 1984; Dunn & Dunn, 1972, 1979; Dunn, Dunn, & Price, 1975, 1978, 1979; Price, Dunn, & Sanders, 1980; as cited in Reid, 1987). The dissemination of a new line of instrumentation in the form of the PLSPQ, the SAS, the LCPC, and the PLPS within applied linguistics, purporting to measure the same perceptual modalities as those measured in the earlier work of the Dunns, seems to have occurred despite substantial objections to the viabil-

ity of this measurement, and the predictive power of what is supposedly measured, in the late 1970s and early 1980s.

Kampwirth and Bates (1980) and Tarver and Dawson (1978), in secondary research examining a number of studies, considered interaction effects between modality preference and teaching strategies and found no convincing empirical support for such interactions. Deverensky (1978) was the first to propose that the problem might be operational, arguing that the task of finding sensitive measures of such preferences was difficult. In a sequence of rebuttal and rejoinder a decade later which involved Kavale and Forness (1987; 1990) and R. Dunn (1990), the predictive power of modality preference was once more brought into question. Kavale and Forness (1987) performed a meta-analysis of 39 studies and concluded that the effect size of the interventions was small. Critically for assertions that will later be made in this paper, they also argued that the measurement of modality preference was difficult. This was consistent with observations made by Deverensky 10 years previously.

The point of recounting the above debate² is to demonstrate that even as perceptual learning styles research got off the ground in applied linguistics in the late 1980s, there was an existing rebuttal to answer to concerning the operational viability of such constructs and their predictive power. My own survey of the literature within applied linguistics indicates that, to the best of my knowledge, at no point has any perceptual learning styles researcher engaged directly with these early objections from outside the field. It seems that initiators of perceptual learning styles research within applied linguistics either did not know about the debate or neglected to engage with it. For whatever reason, the omission was significant because the contours of perceptual learning styles research within applied linguistics were fated to retrace those from outside of the field in a notable case of repeating history. A new set of constructs would be offered to the applied linguistics community along with companion instrumentation. No convincing evidence of predictive power would then be demonstrated for these constructs. And later, the operational viability of these constructs would be drawn into question.

Perceptual Learning Styles Research in Applied Linguistics and Psychometric Weakness

Reid (1987), as stated above, is widely considered a seminal paper, if not the seminal paper, in perceptual learning styles research within applied linguistics in that it introduced the PLSPQ to applied linguistics literature,

and its presence ever since has been significant in this literature (Bowman, 1996; Hyland, 1993; Isemonger & Sheppard, 2003; Kim, 2001; Melton, 1990; O'Donoghue et al., 2001; Peacock, 2001; Reid, 1998a, 2000; Rossi-Le, 1995; Siew Luan & Ngoh, 2006; Stebbins, 1995; Yu-rong, 2007). A search on the Internet will also reveal frequent use in unpublished postgraduate theses, symposia, and other forums. This does not include use in action research and classroom practice, the prevalence of which is difficult to determine empirically.

The PLSPQ is claimed to measure six constructs relating to learning preference, four of which are perceptual (Visual, Auditory, Kinesthetic, and Tactile) and two of which are social (Group and Individual). Reid used the instrument in 1987 without validation of scores. A subsequent (1990) paper by Reid, in the same journal, dealt with validation by reporting Cronbach's alphas, but these were not fully reported for all constructs on the final version of the instrument seen in the 1987 study and still in use today. The 1990 paper also chronicled a problematic development process for the instrument involving construct-related difficulties that have never been overcome.

The pervasive and continuing use of the PLSPQ presents a case study in the cautioning issued by Wilkinson and the American Psychological Association (APA) Task Force on Statistical Inference (1999, p. 596) concerning the tendency for defective measures to remain in use once they have entered the literature. Reid's (1990) article provided an open and forthcoming account of the problematic development of the instrument and the shortcuts taken that should have arrested its further use pending revision and demonstration of the capacity to generate valid scores. Unfortunately, the instrument had already gained momentum in the literature and its use persisted and even grew. The most significant challenge to the capacity of the instrument to produce psychometrically valid scores came as late as 14 years later (Wintergerst, DeCapua, & Itzen, 2001) in a study that employed Exploratory Factor Analysis (EFA) and Cronbach's alphas as diagnostics—although Itzen (1995) had examined the issue earlier in a dissertation that gained little exposure and that employed the additional method of Confirmatory Factor Analysis (CFA). The Cronbach's alphas from all of these studies are available for inspection in Table 1. Also included is the more recent study by Isemonger and Sheppard (2007) which employed both the methods of EFA and CFA in addition to reporting alphas.

Table 1. Comparative Alphas for Past Studies of Scores Generated by the PLSPQ

Authors	Language	Sample	V	A	K	T	G	I
Itzen (1995)	English	92 NSs	.47	.46	.66	.76	.88	.78
		126 NNSs	.54	.56	.63	.72	.87	.80
Wintergerst et al. (2001)	English	100	.37	.39	.69	.59	.87	.75
Isemonger and Sheppard (2007)	Korean	691	.37	.39	.76	.67	.83	.84
Strength			Weak	Moderate	Strong			

NS = Native Speaker; NNS = Nonnative Speaker

V = Visual, A = Auditory, K = Kinesthetic, T = Tactile, G = Group, I = Individual

The following observations are pertinent. The part of the PLSPQ that measures perceptual modality performs marginally in the case of the Kinesthetic and Tactile scales and poorly for the Visual and Auditory scales. Assuming Nunnally and Bernstein's (1994) criterion of .7 for scale reliability,³ the Visual and Auditory scales' alphas are inadequate. In terms of this same criterion, the Kinesthetic and Tactile scales are marginal in performance. Furthermore, the studies conducted by Wintergerst et al. (2001) and Isemonger and Sheppard (2007) present results which threaten the claim for the discriminant validity of the Kinesthetic and Tactile scales. EFAs conducted by Isemonger and Sheppard and Wintergerst et al. failed to reduce to simple structure in line with the scoring model offered for the instrument. Finally, CFAs⁴ conducted by Isemonger and Sheppard and Itzen (1995) failed to confirm Reid's six-scale model.

Turning from the PLSPQ to two of the other perceptual leaning style instruments, the LCPC (Learning Channel Preference Checklist; O'Brien, 1990, 2002) and the SAS (Style Analysis Survey; Oxford, 1993a, 1993b), the omission to demonstrate valid scores in sound psychometric studies was a concomitant feature of the emergence of these instruments in the literature. In view of the problematic psychometrics of scores generated by the PLSPQ in studies cited above, Isemonger and Watanabe (2007) conducted research into the psychometrics of scores generated by the perceptual component of the SAS. The stated goal of the research was to assess whether operational problems were specific to the PLSPQ or a possible feature of the general line of instrumentation. Results for scores on the SAS were poor. Values for

Cronbach's alpha were as follows: Visual, .69; Auditory, .56; and Hands-On, .58. Given that these are 10-item scales⁵ one would expect the alpha values to be considerably higher. Furthermore, in the same study, the model hypothesized by Oxford's design and scoring regime for the instrument was not confirmed in a CFA, and an EFA indicated problems with operationalizations including labels for constructs that were wider than the operational bandwidth of the construct. This study extended the problematic from the PLSPQ to other instruments measuring such constructs. Another by Isemonger (2008) examining the psychometrics of scores generated by the LCPC has brought these instruments further into question. Cronbach's alphas in this study were as follows: Visual, .52; Auditory, .42; and Haptic, .51. Again, given that these are 12-item scales, one would expect the value for alpha to be considerably higher. The model offered in the scoring regime for the instrument was not confirmed in a CFA. I am not aware of any prior research that had comprehensively examined the psychometrics of a set of scores generated by these instruments, using either EFA or CFA as the method.

While not as widely used as the PLSPQ, the LCPC and the SAS have been employed in both research and applied practice. The LCPC is commercially available (Specific Diagnostics Inc.) and its emergence in applied linguistics literature came through Reid (1995) in a regularly cited book directed at the practitioner. It has appeared in research (Hughes, 2001; Oxford, Young, Ito, & Sumrall, 1993). The first version received a negative review in the BUROS Institute of Mental Measurements' publication (Deaton, 1992). My own research (Isemonger, 2008) used the modified 2002 version. The SAS entered applied linguistics literature through Reid's (1995; 1998b) books—the 1998 book also being directed at the practitioner. It is also available for students to self-administer on the University of Alabama's College of Arts and Science's website (Oxford, 1993b). The instrument has entered the literature pertaining to Japanese as a foreign language (Ehara, 1998) and has also found use in recent research (Henry-Vega, 2004; Psaltou-Joycey & Kantaridou, 2011). Again, use in action research and classroom practice is hard to assess empirically but, given entrance into the literature, such usage is presumed. With regard to perceptual learning styles, therefore, the current situation within applied linguistics is one of considerable concern. Instruments have entered the literature and been used over a period of 20 to 25 years without sufficient documentation of the development process, and have even been launched without provision of minimal indexes of reliability such as Cronbach's alpha. None of the instruments (PLSPQ, SAS, nor LCPC) entered the literature accompanied by results from an EFA or CFA

to justify the model (implicitly hypothesized in the scoring regime for each respective instrument) at launch time.⁶ Furthermore, the limited research that has been done by independent researchers, after the introduction of the instruments, provides no reassurance. In fact, such research exacerbates the doubt.

Perceptual Learning Styles and Evidence of Predictive Power

As explained above, one of the central objections to perceptual learning styles constructs prior to their emergence as an area of interest within applied linguistics was that they lacked predictive power (Deverensky, 1978; Dunn, 1990; Kampwirth & Bates, 1980; Kavale & Forness, 1987, 1990; Tarver & Dawson, 1978). Given the strength of the objections, it would have been appropriate for applied linguistics to engage with this issue directly and from the outset—of course, after establishing good psychometrics which is necessarily prior. However, what we have seen is a research trajectory that is almost completely descriptive in nature. Very few studies have attempted to demonstrate the predictive power of these constructs in terms of learning outcome, and those that have introduced an achievement criterion remain correlational studies and, anyway, have pointed to little correlation. For example, Ehrman and Oxford (1995) found no significant correlations between learning style constructs measured by the Learning Styles Profile (LSP; Keefe, Monk, Letteri, Languis, & Dunn, 1989), a comprehensive learning styles instrument that includes perceptual constructs, and speaking and reading proficiency.

Bailey, Onwuegbuzie, and Daley (2000) state with regard to learning styles in general, rather than perceptual learning styles specifically

There appears to be a gap in recent research between logical analyses of the importance of learning styles for foreign language learning and statistical confirmation of learning style preference as a direct measure of foreign language achievement. (p. 128)⁷

In the same study, and in an effort to address the deficit, Bailey et al. (2000) examined the role of a range of learning styles, including perceptual learning styles, in predicting foreign language achievement. The instrument of choice for these authors was the PEPS (Productivity Environmental Preference Survey), referred to above and a progenitor of the PLSPQ, which measures learning style in four major areas: preferences for environmental stimuli

(sound, light, etc); emotional stimuli; sociological stimuli; and physical stimuli (the category into which the auditory, visual, and kinesthetic modes of perception fall). Reliability for the subscales used in the study could not be assessed because scoring was done by the owners of the instrument (Bailey et al., 2000)—an extraordinary limitation to place on an instrument by any author or publisher. Foreign language achievement was measured using standardized course averages to accommodate for differences in teacher characteristics. Multiple regression (All Possible Subsets: APS) and correlation analysis were conducted with the learning styles constructs functioning as the independent variable and the achievement scores as the dependent variable. Of the perceptual learning styles constructs represented in the instrument, only kinesthetic preference correlated significantly with the dependent measure (achievement scores) accounting for approximately 4% of the variance. The auditory and visual constructs correlated very weakly with the dependent measure. In the APS multiple regression model, kinesthetic preference featured along with the design, responsibility, and mobility constructs. The overall model accounted for 15.1% of the variance.

In a less well-known study, Thomas, Cox, and Kojima (2000) obtained similar results. Achievement measures were TOEIC (Test of English for International Communication) scores and class grades, and the learning styles instrument was the PLSPQ. No correlation between styles and TOEIC scores was found, and only the kinesthetic scale (similar to the hands-on scale in the SAS) showed some correlation with course grades in practical, skills-based courses (the achievement measure). The kinesthetic (hands-on) construct is difficult to locate as a perceptual construct because much of the behavior associated with such a construct involves integrated perception, although tactile perception might be more dominant in the construct. Nonetheless, the two most salient modalities of perception, auditory and visual, do not feature in the results.

These results are modest to poor overall, and most importantly, provide no evidence for the predictive power of preferences for the most obvious perceptual constructs, namely, the visual and auditory modalities. It is also notable that these are the constructs which have proved the most problematic from a psychometric point of view (Table 1). Bailey et al. (2000), in accounting for the weak performance of learning styles constructs within their study, commented on instrumentation issues in their conclusion. While made in the context of learning styles in general, these comments are just as applicable to the subset of perceptual learning styles. The two most pertinent are a call for more situation-specific instruments, and an endorse-

ment of Skehan's (1991) recommendation that it may be necessary to move beyond questionnaires—a recommendation that has still not seen adoption; see more detailed discussion below. Suffice to state at this point that issues of instrumentation necessarily precede issues of demonstrating predictive power empirically, and that instrumentation issues may very well go to the heart of the disjunction between the intuitive appeal of perceptual learning styles constructs and the lack of empirical support for their usefulness. The track record of perceptual learning styles instrumentation is not good, and I argue that editorial oversight in enforcing the counsel of the APA, counsel specifically offered by Wilkinson and the APA Task Force on Statistical Inference (1999) and relating to the entry of defective measures into the literature, is critical.

Some Recommendations Related to Editorial Oversight on Entry of Instrumentation Into the Literature

A broader, noteworthy implication of the above analysis is that weak instruments have gained traction and persistence within applied linguistics contra the above-mentioned warnings of Wilkinson and the APA Task Force on Statistical Inference (1999). One lesson which should be learned from this state of affairs is that editorial oversight should be particularly austere with respect to new instruments entering the literature—particularly when these instruments are recommended to practitioners for classroom diagnostics. It is considerably harder to call attention to the limitations of an instrument after entry into the literature than it is to simply prevent entry right from the beginning. Once researchers are using an instrument, other researchers tend to assume that the instrument must be credible due to a mainstream effect. This is human behavior.

More specifically, and ideally,⁸ I argue that new instruments should enter the literature via a dedicated psychometric study. By “dedicated,” it is meant that the entire study focuses in a comprehensive way on the psychometric properties of scores produced by the instrument for an intended population, and the psychometric issue is not something that is simply auxiliary to other research goals.⁹ In such a study, the standard quotation of Cronbach's alphas (an index that has come to be seen erroneously as a panacea for assessing reliability within much of applied linguistics) is not sufficient for reasons that will be outlined below. Insofar as demonstrating unidimensionality of scores is one important component of such a comprehensive study, CFA (with further discussion below) is advocated as a powerful tool in satisfying this requirement. It is important that my assertion of the negative case, that

alpha is not a panacea, not be read as implying a corresponding positive case that CFA is a panacea. CFA is a subset of structural equation modeling (SEM) and refers to the measurement model as opposed to the structural model. In other words, CFA is concerned with the manner in which observable measures reduce to purported underlying constructs rather than with how these constructs relate to each other. If a dedicated measurement paper is not within the remit of a particular applied linguistics journal, then the editor, assisted by the editorial board, should ensure that such a paper has been published or at least gone to press in a journal receptive to such studies before allowing the instrument to enter the literature for any other purpose via their own journal. Omission in this respect may involve the journal in being the conduit for an inadequate instrument gaining traction in the research community, and importantly contra the recommendations of Wilkinson and the APA Task Force on Statistical Inference (1999).

It was stated above that it is not sufficient to simply report Cronbach's alphas in the kind of dedicated psychometric study suggested. It needs to be emphasized that Cronbach's alpha is a minimal index and it is far from the state of the art. In fact, recent argumentation has drawn attention to its limitations and called for its replacement as a routine index of reliability (Bentler, 2009; Green & Yang, 2009a, 2009b; Revelle & Zinbarg, 2009; Sijtsma, 2009a, 2009b). Cronbach's alpha has historically been seen as useful in assessing reliability of scores in ongoing uses of an instrument for research purposes after a dedicated and comprehensive psychometric study for the instrument at launch time. This routine engagement with the reliability of scores is advisable because prior generation of reliable scores in one population does not guarantee scores with similar reliability in another population using the same instrument, and alpha has been seen as an easy and convenient reliability index to compute in dealing with this (though its continued suitability is a matter of objection as the citations above indicate). Wilkinson and the APA Task Force on Statistical Inference (1999) have emphasized that reliability is a property of scores and not instruments, and for this reason indexes of reliability should be reported for any new set of scores even when the purpose of a particular study is not psychometric.

The first limitation with respect to alpha has been pointed out by Cortina (1993), a study cited by Sijtsma (2009a), who drew on a Monte Carlo study¹⁰ conducted by Green, Lissitz, and Mulaik (1977)¹¹ which was also a study cited by Sijtsma. This limitation is that Cronbach's alpha is partly a function of the number of items in a scale. In illustrating this point, Cortina compared the meaning of standardized alpha = .80 for two hypothetical scales of 3 and 10 items, respectively. Given this alpha level, the inter-item correlation for

the 10-item scale is .28 and for the 3-item scale .57. As Cortina pointed out, these inter-item correlations are strikingly different. This means that scales with a larger numbers of items are predisposed towards higher alpha levels (assuming inter-item correlations are constant) and that alpha levels should always be interpreted critically. More specifically, this has two implications. First, frequently quoted cut-off criteria for alpha, the most common one being Nunnally and Bernstein's (1994) criterion of .70, need to be considered critically when being applied, assuming the advisability of using alpha which is in dispute. Such criteria are useful rules-of-thumb but are not infallible. Second, the index should be understood and interpreted rather than simply reported. It is a diagnostic with limitations and not a panacea for detecting reliability.

Another limitation of alpha is that it is not a good measure of unidimensionality¹² which is a property of scores generated by a scale that needs to be empirically demonstrated if the scale is to be confidently interpreted. Alpha is more precisely a measure of reliability, which is not an equivalent property to unidimensionality. Cortina (1993) stated the following with regard to the psychological literature at the time, and there is a strong case that this pertains to much of applied linguistics literature today:

The problem is that, just as the psychological literature reflects no clear understanding of the extent to which alpha is affected by the number of items, so does it reflect no clear understanding of the extent to which alpha is affected by dimensionality. (p. 101)

The intuitive assumption upon which many researchers are currently operating within aspects of applied linguistics is that if alpha is high, then this is evidence of unidimensionality for scores generated by the scale in question. This is not the case, and Cortina (in engaging with psychological literature) again drew on the Monte Carlo study of Green et al. (1977) to demonstrate this. Miller (1995) put the issue succinctly in stating that the proper use of alpha assumes unidimensionality rather than demonstrating it. It is entirely possible to arrive at a high alpha coefficient from a multidimensional scale. This is often readily apparent in the errant manner in which alpha is often reported with regard to instruments within applied linguistics literature—and here the review extends the claim to areas beyond learning styles. For example, Oxford (1996) reports very high alphas (above .90) for the Strategy Inventory for Language Learning (SILL) which comprises six subscales—and which has found significant use within contemporary ap-

plied linguistics research. In this case, high alphas are reported for an entire instrument that is not unidimensional and is, in fact, multidimensional by design.¹³ And on a further note, the value for alpha in these quotations for entire multiscale instruments (involving many items) is often very high, and this can be expected given alpha's positive bias for number of items. Researchers would be better served by seeing alpha as assuming unidimensionality rather than demonstrating it; that is, if the index continues to be used in spite of the calls for its replacement.

With regard to this paper's advocacy of CFA as an important method for attending to the issue of unidimensionality of scores in comprehensive psychometric studies to attend the launch of new instruments, an influential article by Gerbing and Anderson (1988) made a case that CFA is the only method properly equipped to assess the unidimensionality of scales. I do not seek to endorse this strong version of the advocacy of CFA and its exclusiveness for testing unidimensionality, but do wish to draw the attention of the reader to its usefulness as one powerful instrument for this purpose. The goals of any researcher releasing new instrumentation into applied linguistics literature should, amongst other things, include demonstrating that scales making up the instrument generate scores that are indeed unidimensional. Subscales that produce multidimensional scores should be considered suspect in terms of validity because they cannot obviously, and necessarily, be interpreted in terms of the single label that semantically characterizes each scale.

Why CFA is powerful with respect to demonstrating unidimensionality is helpfully understood against the more frequently encountered method of EFA, which also engages with dimensionality, but which does not offer the same prospects for a direct test of a hypothesized unidimensional measurement model for an instrument. This applies whether the model in question is explicitly hypothesized by the author or implicitly hypothesized by virtue of the scoring regime offered along with the instrument. In EFA, as Thompson (2004) points out, simple structure is arrived at through a linear sequence of decisions which include:

1. Which matrix of association coefficients should be analyzed?
2. How many factors should be extracted?
3. Which method should be used to extract the factors?
4. How should the factors be rotated?
5. How should factor scores be computed if factor scores are of interest?
(p. 27)

As Thompson and Daniel (1996, p. 204), in an earlier contribution, point out, an expected model either emerges out of this sequence or it does not, and rival models are not tested. Also, the decision of how many factors to extract, while not arbitrary, is attended by difficulties of determinacy. For example, the frequently-used, eigenvalue-greater-than-one rule (Guttman, 1954; Kaiser, 1961) can overestimate the number of factors (Zwick & Velicer, 1986) and inspection of a scree plot (Cattell, 1966) is perceptually subjective. In CFA, the researcher approaches the data set with an a priori model which is tested directly against the data. The unidimensional model is specified so that observables (measured items on the instrument itself) load only on the factor they are hypothesized to indicate and not on other factors and this is unlike EFA, wherein observables can indicate all factors. Adjudication of the fit of the a priori model is conventionally undertaken using a variety of indexes such as the standardized root mean square residual (SRMSR), the root mean square error of approximation (RMSEA), the comparative fit index (CFI) and the Tucker-Lewis index (TLI). There are two important points to note with regard to the use of these indexes. First, the fact that a variety is used provides for triangulation of the decision as to whether the model is satisfactory or not. Second, the cut-off criteria conventionally used to adjudicate model fit (see, for example, Hu & Bentler, 1999) are empirically derived to minimize both Type I and Type II error. This assists, in an evidential way, with the problem of determinacy in adjudicating model fit. Finally, and unlike EFA, the testing of an a priori model can be conducted in the context of testing rival models. Even if the a priori model fits satisfactorily, it is, therefore, possible to test whether other plausible models fit better. This kind of analytical leverage exceeds that available to EFA, and the explanation here is cursory (for further explanation and understanding see Byrne, 2001, 2005; Schmitt, 2011; Thompson, 2004).

Turning away from the statistical concerns referred to above, one of the issues that has emerged in perceptual learning styles research is foreign language instrumentation. A precedent for administration of perceptual learning styles instruments in a language foreign to the respondents was set by Reid (1987), who used an English-language version for NNSs. This precedent has been followed in a number of studies (Bowman, 1996; Peacock, 2001; Rossi-Le, 1995; Stebbins, 1995), but not all studies, leading up to the present. As recently as 2005, DeCapua and Wintergerst have defended their use of English-language instrumentation for NNSs in a line of research using the PLSPQ (or new version of it) that includes four studies (DeCapua & Wintergerst, 2005; Wintergerst & DeCapua, 2001; Wintergerst et al., 2002,

2003). DeCapua and Wintergerst (2005) cited Eliason (1995) who cited Melton (1990) and Inclan (1986) in making an unassertive case for foreign language instrumentation. Eliason reported that Melton's study found no significant difference in scores (using ANOVA and Tukey's Multiple Comparisons of Means) for individuals when the instrument was administered in both English and Chinese. The reasoning here for the case of equivalence of measurement was weak. The psychometric case for different language versions of an instrument measuring equivalently has to be made on the specific merits of the scores generated by each language version of the instrument, in the population for which it was translated, using EFA, CFA,¹⁴ and other methods related to psychometrics as the methodology, and not through the use of standard inferential statistics involving samples of convenience to see if one arrives at a nonsignificant result. This approach taken by Melton was unorthodox,¹⁵ was not the main part of his study, and was no foundation for a precedent.

The citation of these two particular studies (Inclan, 1986; Melton, 1990) by Eliason (1995) and DeCapua and Wintergerst's subsequent (2005) citation of Eliason indicate a view on the part of these authors that some kind of empirical case can be made for the use of language versions of instruments that are foreign to the respondent. However, there is a far stronger case that instruments in language versions foreign to the respondent become progressively more untenable as the language competence of the respondent decreases. An English-language version of the PLSPQ, or any other instrument for that matter, might seem to function with students above a certain threshold (which respondents in DeCapua and Wintergerst's study might have been), but any instructor with the experience of facing classes of beginners will be unconvinced by foreign language instrumentation. An empirical case hardly needs to be made for this. If researchers want instrumentation to accommodate all levels of students, they should translate instruments into the native language of the intended respondents. This is a criterion which needs to be explicitly adopted in any editorial review processes within applied linguistics journals.

It is also notable that this trend toward using foreign language instrumentation within aspects of applied linguistics, which I would argue is significant rather than pervasive, is contrary to the Test Adaptation Guidelines of the International Test Commission (ITC, 2001). The Commission is extensively engaged with establishing statistical and methodological guidelines for adapting tests across languages and cultures, and the precedent for foreign language instrumentation set in the perceptual learning styles research

within applied linguistics directly contradicts this. Any exercise of editorial prerogative to exclude foreign language instrumentation from applied linguistics research and journals would be consistent with the guidelines of the ITC.¹⁶

A Revised PLSPQ and the Future of Instrumentation in Perceptual Learning Styles

As part of this critique of perceptual learning styles research, I now turn to the Learning Styles Inventory (LSI), which is important because it is a recent incarnation of the PLSPQ which may have an impact on the future of learning styles research. The instrument represents an effort to revise the PLSPQ in view of the evident problems, but I argue that the approach, nonetheless, inherits these problems and that the new instrument is premised on a misunderstanding of the concept of reliability via its use of repetitive items.

The LSI is essentially a residual-item PLSPQ under a new hypothesized structure after a diagnostic EFA published by Wintergerst et al. (2001). It has appeared in four other studies (DeCapua & Wintergerst, 2005; Wintergerst & DeCapua, 2001; Wintergerst et al., 2002, 2003). While it represents the latest generation of instrumentation in perceptual learning styles research, the LSI is conspicuous for having abandoned the most salient perceptual constructs (visual and auditory perception) in this process of revision. The hypothesized three-construct model includes three explicit scales: Group Activity Orientation (GAO), Individual Activity Orientation (IAO), and Project Orientation (PO). These scales incorporate most, but not all, of the original items from the PLSPQ.

I question the appropriateness of retaining the items making up the PLSPQ under a different hypothesized structure. The theoretical rationale for the new hypothesized structure for the PLSPQ, in the form of the constructs in the LSI cited above, came after the fact of the EFA rather than before it. A better and more scientifically credible procedure, having established that the PLSPQ generates psychometrically weak scores, would be to proceed with an entirely new instrument based on a priori theoretical reasoning. The hypothesized constructs emerging from such a theoretical rationale should employ, and benefit from, a large and diverse exploratory set of items targeted at the hypothesized constructs rather than the limited set of items in the PLSPQ. The items in the PLSPQ were designed with a different theoretical rationale and hypothesized structure in mind and represent a

post-reduction set of items (reduced by Reid in the original development of the PLSPQ). Furthermore, they were simplified, often to the point of repetitiveness and paraphrasing one another (an issue covered in more detail below), to accommodate the nonnative speaker (Reid, 1990)—a process which belies the view that foreign language instrumentation is not really problematic.

I argue that the line of research being conducted by DeCapua and Wintergerst (2005) with the LSI inherits and perpetuates the problems of the past. The PLSPQ is a very problematic instrument by the author's own admission (Reid, 1990), and it is clearly time to abandon this artifact of applied linguistics' psychometric legacy.

Turning to the issue of repetitive questions in the LSI (inherited from the PLSPQ), DeCapua and Wintergerst (2005) state the following with respect to their use of the instrument in one study:

Both in the class discussions and in the interviews, informants pointed out repeatedly the repetition among the statements. The students were very much aware that the same questions were asked in different ways on the LSI and questioned why. Indeed, in the class discussion several students mentioned that they had thought this was some sort of mistake. Even though the instructor pointed out that stating the same thing in more than one way is a way of ascertaining whether there is consistency in responses, the students still felt that this was a weakness of the LSI. (p. 9)

The informants rather than the instructor have the stronger case here. The purpose of having more than one item measuring a construct is not to ascertain consistency in responses if that means repeating the same question to some degree or another. The purpose is rather to allow for a more diverse and exhaustive operational expression of the underlying construct. If a scale comprises many diverse items, the idiosyncrasies of each specific item in measuring the respective construct will be averaged out. Statements should not be repeated or paraphrased to see if there is consistency in response, because the consistency or stability of the item is with respect to the measurement of the underlying construct and not the item itself. Repeating the item simply repeats the idiosyncrasy of that specific item. Statements that are virtual paraphrases of each other often inter-correlate highly and produce higher alphas but the researcher should hardly be surprised by this. If you ask someone the same question twice, you should not be surprised if you

get the same answer. This use of paraphrases produces measured constructs that have been referred to by Kline (1994) as “bloated specifics.” In addition, the cost of this artificially high internal consistency is that construct bandwidth is sacrificed because there is really only one operational expression of the construct repeated many times. The goal should be to reach optimum levels of both internal consistency and construct bandwidth, and this cannot be done with the items in the PLSPQ, which are flawed as a result of their repetitiveness.

The Future of Perceptual Learning Styles

Perceptual learning styles represent a special case for measurement due to their resistance to operationalization—especially through self-report. Self-report assumes metacognitive awareness of what is being reported and this assumption is not always satisfied. Most of what we perceive as humans is an integrated perceptual experience, and it may be quite difficult for respondents to distil this integrated experience into pure perceptual modalities and make an authentic judgment on which one they favor.

The view that perceptual preference is resistant to operationalization is supported by early research within education (Deverensky, 1978; Dunn, 1990; Kampwirth & Bates, 1980; Kavale & Forness, 1987, 1990; Tarver & Dawson, 1978) and a stream of research experience within applied linguistics that is very much a facsimile of this earlier educational research. There is also an interesting reprise of this issue, again from outside the field of applied linguistics, where operational issues and the usefulness of the constructs have been empirically challenged (Kratzig & Arbuthnott, 2006). Interestingly, this resistance to operationalization is a critical feature of the LSI’s emergence out of the PLSPQ that is not explicitly referred to by DeCapua and Wintergerst (2005). The PLSPQ was originally a perceptual learning styles instrument (Visual, Auditory, Kinesthetic, and Tactile) including two nonperceptual constructs (Group and Individual). The LSI, as a revised version of the original instrument, is a nonperceptual learning styles instrument including, essentially, the original Group and Individual constructs (now the GAO and IAO) and the new Project Orientation (PO) construct, which is an amalgam of the Tactile and Kinesthetic items. This means that in revising the PLSPQ, the notion of perceptual learning styles in the instrument has effectively been abandoned. This fact is very revealing and is not surprising because research into the PLSPQ has consistently shown that the perceptual constructs are weak and the nonperceptual constructs strong (Isemonger & Sheppard, 2007; Itzen, 1995; Wintergerst et al.,

2001). It is clear that perceptual constructs are not viably measured through self-report.

Bailey et al. (2000) endorsed Skehan's (1991) call to move beyond questionnaires with respect to learning styles. With regard to the preference for perceptual learning styles, specifically, this is particularly the case. None of the instruments so far have succeeded in producing valid scores. This recommendation, however, has different implications for practitioners and researchers. For the researcher, there may well be methods within a laboratory setting to reliably determine preferences for perceptual modality. However, for the practitioner, the self-report method is pervasive because it is convenient, efficient, and unobtrusive. If preference for perceptual learning styles cannot reliably be determined using this method, then there may well be no way to effectively measure such styles in the classroom. Given that the literature has yet to credibly demonstrate the predictive power of perceptual learning styles constructs in achievement terms, I argue that practitioners need not lose any sleep over this. There are other instruments in other areas of individual differences that present far better prospects for pedagogical intervention.

Conclusion

It is important to reiterate that editorial oversight in the area of psychometrics is critical to avoiding the problems that have attended the perceptual learning styles research trajectory within applied linguistics. Admitting questionable instruments into the literature is something that must be vigilantly guarded against. I do recognize the difficulties editors face in having to work with the submissions they have rather than the submissions they would desire, and perhaps this also raises the importance of appropriate training in postgraduate applied linguistics programs. What is clearly arguable is that the burden of improving oversight with respect to these issues falls first to leading journals which set the precedent for the field and enjoy greater freedom in their editorial decisions for having more and better submissions.

The conclusions from this critical review are also significant for teachers and practitioners, who have limited time for diagnostics related to individual differences in learning outcome—they do after all have to get down to the core task which is to teach language. There are instruments in other areas of individual differences which have less questionable psychometrics and a far better track record in predicting learning outcome (e.g., anxiety and motiva-

tion), though I am not at all claiming that deficiencies I have pointed out are exclusive to the perceptual learning styles domain in applied linguistics. I would not have cited the case of perceptual learning styles as instructive if the deficiencies were unique. For teachers and practitioners who remain concerned about perceptual preferences, it is arguable that paying attention to offering a good multimodality class covers the proclivities of all groups of students minus the fuss of having to diagnose what preference each student has. Furthermore, it is arguable that all groups are better served by an integrated, multimodality learning experience than a singular modality experience which caters to the preference of a particular group—with all the attendant problems of coping with the groups that might be marginalized by such disadvantageous matching. Assuming that preferences for perceptual modality do exist, even if difficult to detect, there is no reason at all to go on to assume that a visually oriented student will perform better in a visually delivered class than in a multimodality class. In fact, it could be predicted that all perceptual orientations will perform better in an integrated multimodality class, and this prediction might be tested in future research if such orientations or preferences can be reliably detected.

Finally, practitioners should not assume that psychometric instruments (measuring any kinds of constructs) which come out of published literature produce valid scores simply because such instruments have been published and have been used by others. Practitioners should check for good psychometric studies that thoroughly question an instrument, and should not simply be satisfied with uncritical quotations of the value for Cronbach's alpha. High values for alpha quoted for an entire instrument that is multidimensional by design are not very helpful because interpretation usually occurs at the level of each dimension or scale, and anyway alpha does not effectively demonstrate unidimensionality, which is a critical property for score interpretation. Furthermore, high values for alpha on scales which contain highly paraphrased items should be treated with considerable skepticism. In terms of all of the above, the line of instrumentation within applied linguistics measuring perceptual learning styles is questionable and practitioners should avoid using these diagnostics in class until the instruments are shown to be capable of generating valid scores, and until there is serious evidence for the predictive power of the constructs they attempt to measure in terms of learning outcome. There is an opportunity cost in running diagnostics in language class, and this cost is lost time for the core business of language teaching. This opportunity cost should only be borne if the benefits from the diagnostics have been empirically demonstrated to be significant.

Notes

1. I also neglected this work as a new researcher entering this area (Isemonger & Sheppard, 2003). In the same study in 2003, I also failed to observe some of the recommendations made in this paper. I have also pointed out these failures once before (5 years ago) in Isemonger and Sheppard (2007). In addition, some of the claims about perceptual learning styles in the 2003 paper illustrate how much I have changed my mind since that time.
2. I have recounted this history of learning styles controversy which occurred outside of applied linguistics once before in an article in the *Journal of Psychoeducational Assessment* (Isemonger, 2008).
3. In a meta-study drawing on a sample of 696 tests appearing in the APA-published Directory of Unpublished Experimental Mental Measures (Hogan, Benjamin, & Brezinski, 2000), 85% of reported alpha coefficients met this criterion of .70. However, as will be argued later in this paper, Cronbach's alpha is an index that should be interpreted. The number of items on a scale "biases" (positively) the value derived for the index as true score variance builds up faster than error variance as items are added. In the case of the results presented for the Auditory and Visual Scales on the PLSPQ, there is little need for interpretation of alpha with respect to this bias since the value for alpha is so low.
4. CFA is a more sophisticated and appropriate verificational tool than EFA and also allows for the direct testing of the superiority of rival structures for an instrument. There is more discussion on this later in the paper.
5. Again, and this will be explained in more detail further on in the paper, Cronbach's alpha is biased (positively) by the number of items in a scale: 10 to 12 items would be considered a fairly high number of items for a scale leading to an expectation of higher alphas.
6. CFA is a method that has been adopted more recently than EFA due to the specialized software required for it, and retrospective criticism with regard to the absence of its use for emerging instrumentation should be understood in this context.
7. Bailey, Onwuegbuzie, and Daley do cite a study conducted by Oxford et al. (1993) which dealt with perceptual learning styles using the LCPC and which claimed that the Visual construct was "more predictive of Japanese language achievement" (p. 367) than the Auditory and Kinesthetic constructs. However, no achievement data is presented in the Oxford et al. study, nor were the statistical methods for arriving at this conclusion reported. No models were presented indicating the respective amounts

of variance accounted for by these constructs vis-à-vis other constructs examined in the same study including motivational and learning-strategy constructs. Furthermore, alphas for the LCPC subscales were poor given that all fell well below .70 on 12-item scales. Bailey, Onwuegbuzie, and Daley also review some other studies that consider the predictive power of learning styles, but the studies reviewed examine nonperceptual learning styles constructs.

8. I say “ideally” here because journal editors deal with the reality that the standard of current practices for the submissions they receive (which is of course highly contingent on the stature of the journal) may not necessarily be conducive to enforcing the best practice. Ultimately, the burden of implementation should fall first on the leading journals in the expectation that methodological practices in such journals set the precedent and cascade down to journals with harder choices to make at press time.
9. If the instrument produces valid scores in the first dedicated study, then this should also be followed up with further similar studies in all populations where it is intended for use. Evidence in this regard is a cumulative process.
10. A Monte Carlo study is a study in which certain known properties are built into a data set.
11. I realize that some of these citations may seem dated, but the uncritical manner in which alpha is used as an index of reliability within applied linguistics indicates that their central point has yet to be taken on board. More recent citations of these older studies by Sijtsma indicate that this is also the case in other research fields, and that these older citations are still pertinent.
12. In simple terms, unidimensionality refers to the property that the scale measures quite *exclusively* what the author claims it measures, and does not measure other (possibly unknown) dimensions of behavior that may complicate the interpretation of the scale.
13. The rationale for this reporting of a composite alpha for the entire SILL is not clear. If a second order CFA model hypothesized a superordinate construct which subsumed the six subscales and this model was confirmed and worked as well, then there would be a rationale for quoting alpha for the entire instrument, because interpretation of a composite score for the six subscales would make sense. This might occur with constructs such as anxiety where different facets of anxiety constitute subcomponents of a more generalized anxiety disposition. I do not see

how it occurs in the case of the SILL, and certainly no case has been made for a superordinate and interpretable construct for the overall SILL. Alpha is only meaningful at the level at which interpretation takes place—and unidimensionality should have been previously demonstrated at this level for scores derived from the population for which the scale is being used.

14. One of the powerful ways to establish equivalency of measurement across translated versions of an instrument (administered in their respective populations) is to use the group analysis capabilities of CFA in what is often referred to as a measurement invariance study. Byrne (2001) offers helpful chapters on this analytical procedure.
15. A new language version of an instrument is essentially a new instrument for a new population. Its equivalence with the original needs to be empirically demonstrated and CFA is an appropriate method for this.
16. Quite obviously, this would not include instrumentation where the foreign language being learned is the subject of the test.

Ian Isemonger is an Associate Professor at Kumamoto University where he teaches in the Department of Communication and Information Studies and in the Graduate School's TESOL master's and doctoral programs.

References

- Bailey, P., Onwuegbuzie, A. J., & Daley, C. E. (2000). Using learning style to predict foreign language achievement at the college level. *System, 28*, 115-133.
- Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74*, 137-143.
- Bowman, A. (1996). ESL students learning style preferences in a multimedia language laboratory: Do students do what they say they do? *University of Hawaii Working Papers in ESL, 15*, 1-31.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS*. London: Lawrence Erlbaum Associates.
- Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment, 85*, 17-32.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104.

- Deaton, W. L. (1992). Buros review of LCPC. In J. J. Kramer & J. C. Conoley (Eds.), *The eleventh mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- DeCapua, A., & Wintergerst, A. C. (2005). Assessing and validating a learning styles instrument. *System, 33*, 1-16.
- Deverensky, J. L. (1978). Modal preferences and strengths: Implications for reading research. *Journal of Reading Behaviour, 10*, 7-23.
- Dunn, R. (1983). Learning style and its relation to exceptionality at both ends of the spectrum. *Exceptional Children, 49*, 496-506.
- Dunn, R. (1984). Learning style: State of the scene. *Theory into Practice, 23*, 10-19.
- Dunn, R. (1990). Bias over substance: A critical analysis of Kavale and Forness' report on modality-based instruction. *Exceptional Children, 56*, 352-356.
- Dunn, R., & Dunn, K. (1972). *Practical approaches to individualizing instruction*. Englewood Cliffs, NJ: Prentice Hall.
- Dunn, R., & Dunn, K. (1979). Learning styles/teaching styles: Should they ... can they ... be matched? *Educational Leadership, 36*, 238-244.
- Dunn, R., Dunn, K., & Price, G. E. (1975). *The Learning Style Inventory*. Lawrence, KS: Price Systems.
- Dunn, R., Dunn, K., & Price, G. E. (1978). *Teaching students through their individual learning styles*. Reston, VA: Reston Publishing.
- Dunn, R., Dunn, K., & Price, G. E. (1979). *The Productivity Environment Preference Survey*. Reston, VA: Reston Publishing.
- Ehara, Y. (1998). Differences in learning styles between Japanese teachers of Japanese and Mexican learners of Japanese. *Journal of the Society for Teaching Japanese as a Foreign Language, 96*, 13-24.
- Ehrman, M., & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal, 79*, 67-89.
- Eliason, P. A. (1995). Difficulties with cross-cultural, learning-styles assessment. In J. M. Reid (Ed.), *Learning styles in the ESL/EFL classroom* (pp. 19-33). Boston, MA: Heinle & Heinle.
- Frank, M., & Hughes, M. (2002). Examining the learning styles of Japanese students at Keiwa College. *Keiwa Gakuen University Research Bulletin, 11*, 73-86.
- Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research, 15*, 186-192.

- Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37*, 827-838.
- Green, S. B., & Yang, Y. (2009a). Commentary on coefficient alpha: A cautionary tale. *Psychometrika, 74*, 121-135.
- Green, S. B., & Yang, Y. (2009b). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika, 74*, 155-167.
- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19*, 149-161.
- Henry-Vega, G. (2004). *Exploratory study on the processing styles and the processing strategies of 2 second language graduate students when reading texts for academic purposes*. University of Cincinnati, Cincinnati, OH.
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement, 60*, 523-531.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.
- Hughes, M. (2001). Examining the learning styles and learning strategies used by Japanese learners of English. *Keiwa Gakuen University Research Bulletin, 10*, 59-87.
- Hyland, K. (1993). Culture and learning: A study of the learning styles of Japanese students. *RELC Journal, 24*, 69-91.
- Inclan, A. F. (1986). The development of the Spanish version of the Myers Briggs type indicator, form G. *Journal of Psychological Type, 11*, 35-46.
- International Test Commission. (2001). *International Test Commission guidelines for test adaptation*. London: Author.
- Isemonger, I. M. (2008). Scores on a Japanese-language version of the Learning Channel Preference Checklist: A questionable instrument within a questionable line of instrumentation. *Journal of Psychoeducational Assessment, 26*, 148-155.
- Isemonger, I. M., & Sheppard, C. (2003). Learning styles. *RELC Journal, 34*, 195-222.
- Isemonger, I. M., & Sheppard, C. (2007). A construct-related validity study on a Korean version of the Perceptual Learning Styles Preference questionnaire. *Educational and Psychological Measurement, 67*, 357-368.

- Isemonger, I. M., & Watanabe, K. (2007). The construct validity of scores on a Japanese version of the perceptual component of the Style Analysis Survey (SAS). *System, 35*, 134-147.
- Itzen, R. (1995). *The dimensionality of learning structures in the Reid Perceptual Learning Style Preference Questionnaire* (Unpublished doctoral dissertation). Graduate College of the University of Illinois at Chicago.
- Kaiser, H. (1961). A note on Guttman's lower bound for the number of common factors. *Multivariate Behavioral Research, 1*, 249-276.
- Kampwirth, T. J., & Bates, M. (1980). Modality preference and teaching method: A review of the research. *Academic Therapy, 15*, 597-605.
- Kavale, K. A., & Forness, S. R. (1987). Substance over style: Assessing the efficacy of modality testing and teaching. *Exceptional Children, 54*, 228-239.
- Kavale, K. A., & Forness, S. R. (1990). Substance over style: A rejoinder to Dunn's animadversions. *Exceptional Children, 56*, 357-361.
- Keefe, J. W., Monk, J. S., Letteri, C. A., Languis, M., & Dunn, R. (1989). *Learning style profile*. Reston, VA: National Association of High School Principals.
- Kelly, C. A. (1998). The learning style preferences of Japanese EFL students. *Gakuen, 697*, 30-36.
- Kim, H. J. (2001). Language learning strategies, learning styles and beliefs about language learning of Korean university students. *Journal of the Pan-Pacific Association of Applied Linguistics, 5*, 31-46.
- Kinsella, K. (1995a). Perceptual Learning Preferences Survey. In J. M. Reid (Ed.), *Learning styles in the ESL/EFL classroom* (pp. 221-238). Boston, MA: Heinle & Heinle.
- Kinsella, K. (1995b). Understanding and empowering diverse learners in the ESL classroom. In J. M. Reid (Ed.), *Learning styles in the ESL/EFL classroom* (pp. 170-193). Boston, MA: Heinle & Heinle.
- Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.
- Kratzig, G. P., & Arbuthnott, K. D. (2006). Perceptual learning styles and learning proficiency: A test of the hypothesis. *Journal of Educational Psychology, 98*, 238-246.
- Melton, C. D. (1990). Bridging the cultural gap: A study of Chinese students' learning style preferences. *RELC Journal, 21*, 29-47.
- 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., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- O'Brien, L. (1990). *Learning Channel Preference Checklist (LCPC)*. Rockville, MD: Specific Diagnostic Services.
- O'Brien, L. (2002). *The Learning Channel Preference Checklist (LCPC)*. Kensington, MD: Specific Diagnostics Services.
- O'Donoghue, R. K., Oyabu, T., & Akiyoshi, R. (2001). An exploratory survey of Japanese EFL students' preferred learning styles. *Bulletin of Fukuoka Dental College*, 28, 9-17.
- Oxford, R. L. (1993a). *Style Analysis Survey (SAS)*. Tuscaloosa, AL: University of Alabama.
- Oxford, R. L. (1993b). *The Style Analysis Survey (SAS) on CCET, University of Alabama*. Retrieved February 2010 from: <http://www.as.ua.edu/nihongo/sas/survey.html>
- Oxford, R. L. (1995). Gender differences in language learning styles: What do they mean? In J. M. Reid (Ed.), *Learning styles in the ESL/EFL classroom* (pp. 34-46). Boston, MA: Heinle & Heinle.
- Oxford, R. L. (1996). Employing a questionnaire to assess the use of language learning strategies. *Applied Language Learning*, 7, 25-45.
- Oxford, R. L., & Anderson, N. J. (1995). A cross cultural view of learning styles. *Language Teaching*, 28, 201-215.
- Oxford, R. L., Young, P. O., Ito, S., & Sumrall, M. (1993). Japanese by satellite: Effects of motivation, language learning styles and strategies, gender, course level, and previous language learning experience on Japanese language achievement. *Foreign Language Annals*, 26, 359-371.
- Peacock, M. (2001). Match or mismatch? Learning styles and teaching styles in EFL. *International Journal of Applied Linguistics*, 11, 1-20.
- Price, G. E., & Dunn, R. (1997). *The learning style inventory: LSI manual*. Lawrence, KS: Price Systems.
- Price, G. E., Dunn, R., & Dunn, K. (1996). *Productivity environmental preference survey: PEPS manual*. Lawrence, KS: Price Systems.
- Price, G. E., Dunn, R., & Sanders, W. (1980). Reading achievement and learning style characteristics. *The Clearing House*, 5, 223-226.
- Psaltou-Joycey, A., & Kantaridou, Z. (2011). Major, minor, and negative learning style preferences of university students. *System*, 39, 103-112.
- Reid, J. M. (1987). The learning style preferences of ESL students. *TESOL Quarterly*, 21, 87-109.

- Reid, J. M. (1990). The dirty laundry of ESL survey research. *TESOL Quarterly*, 24, 323-338.
- Reid, J. M. (1995). *Learning styles in the ESL/EFL classroom*. Boston, MA: Heinle & Heinle.
- Reid, J. M. (1998a). Teachers as perceptual learning styles researchers. In J. M. Reid (Ed.), *Understanding learning styles in the second language classroom* (pp. 15-26). Englewood Cliffs, NJ: Prentice Hall Regents.
- Reid, J. M. (1998b). *Understanding learning styles in the second language classroom*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Reid, J. M. (2000). The perceptual learning style questionnaire. In J. M. Reid (Ed.), *The process of composition* (pp. 328-330). New York: Prentice Hall Regents.
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the GLB: Comments on Sijtsma. *Psychometrika*, 74, 145-154.
- Rossi-Le, L. (1995). Learning styles and strategies in adult immigrant ESL students. In J. M. Reid (Ed.), *Learning styles in the ESL/EFL classroom* (pp. 118-125). Boston, MA: Heinle & Heinle.
- Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 29, 304-321.
- Shen, M. Y. (2010). Effects of perceptual learning style preferences on L2 lexical inferring. *System*, 38, 539-547.
- Siew Luan, T.-K., & Ngho, J. (2006). Innovative modes of continual assessment: Perspectives of undergraduate students. *Reflections on English Language Teaching*, 5, 47-64.
- Sijtsma, K. (2009a). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74, 107-120.
- Sijtsma, K. (2009b). Reliability beyond theory and into practice. *Psychometrika*, 74, 169-173.
- Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition*, 13, 275-298.
- Stebbins, C. (1995). Culture-specific perceptual-learning-style preferences of post secondary students of English as a second language. In J. M. Reid (Ed.), *Learning styles in the ESL/EFL classroom* (pp. 108-117). Boston, MA: Heinle & Heinle.
- Tarver, S. G., & Dawson, M. M. (1978). Modality preference and the teaching of reading: A review. *Journal of Learning Disabilities*, 11, 17-29.

- Thomas, H., Cox, R., & Kojima, T. (2000, March). *Relating preferred learning style to student achievement*. Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages, Vancouver, BC, Canada.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: APA.
- Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*, 197-208.
- Wilkinson, L., & American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.
- Wintergerst, A. C., & DeCapua, A. (2001). Exploring the learning styles of Russian-speaking ESL students. *The CATESOL Journal, 13*, 23-46.
- Wintergerst, A. C., DeCapua, A., & Itzen, R. C. (2001). The construct validity of one learning styles instrument. *System, 29*, 385-403.
- Wintergerst, A. C., DeCapua, A., & Verna, M. A. (2002). An analysis of one learning styles instrument for language students. *TESL Canada Journal, 20*, 16-37.
- Wintergerst, A. C., DeCapua, A., & Verna, M. A. (2003). Conceptualizing learning style modalities for ESL/EFL students. *System, 31*, 85-106.
- Yamashita, S. (1995). The learning style preferences of Japanese returnee students. *ICU Language Research Bulletin, 10*, 59-75.
- Yu-rong, H. (2007). A survey on the learning style preference of Tibetan EFL learners in China. *US-China Foreign Language, 5*, Serial No. 43.
- Zwick, W. R., & Velicer, W. F. (1986). Factors influencing five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.

