Electronic Journal of Research in Educational Psychology

ejrep

25 AÑOS 1993 | 2018 UNIVERSIDAD DE ALMERÍA

Ilustre Colegio Oficial de Psicología DE ANDALUCÍA ORIENTAL

Learning Styles Inventories: an update of Coffield, Moseley, Hall, & Ecclestone's Reliability and Validity Matrix

Francisco José Álvarez-Montero[1], **María Guadalupe Leyva-Cruz** [1], **Flérida Moreno-Alcaraz** [1]

[1] Facultad de Ciencias de la Educación, Universidad Autónoma de Sinaloa

México

Correspondence: Francisco J. Álvarez Montero. Facultad de Ciencias de la Educación, Ave. Cedros y Calle Sauces s/n Fracc. Los Fresnos, C.P. 80034. Culiacán, Sinaloa, México.
E-mail: francisco_alvarez_montero@uas.edu.mx

Resumen

Introducción. Una de las muchas nociones que se han introducido en educación en los últimos 50 años, es la de los estilos de aprendizaje. Esta idea es muy popular y goza de buena aceptación. Sin embargo, desde una perspectiva científica esta noción es muy controversial. En este sentido, Coffield, Moseley, Hall, y Ecclestone, realizaron la primera revisión sistemática sobre el tema, y mostraron que los instrumentos psicométricos de estilos de aprendizaje son no tienen niveles adecuados de fiabilidad y validez. Sin embargo, los resultados de su revisión no han sido actualizados desde 2004. Esta contribución presenta una revisión sistemática de estos aspectos para el periodo 2005-2010, con el objetivo de actualizar la información sobre el tema.

Método. Se examinaron 58 documentos que contenían suficiente información sobre fiabilidad y validez acerca de los siguientes inventarios: el LSI de Kolb, el ILS de Felder, el TSI de Sternberg, los inventarios sensoriales tipo VAK-VAKT-VARK, los inventarios del modelo de Entwistle (ASSIST, LASSI y RASI), el SPQ de Biggs, el LSQ de Honey y Mumford y el Delineador de Estilos de Gregorc (GSD).

Resultados. El análisis de las propiedades psicométricas señala que: (1) más de la mitad de los inventarios escrutados carece de consistencia interna y validez predictiva, (2) alrededor del 31% no tiene fiabilidad test-retest, (3) no hay evidencia acerca de por lo menos un criterio de fiabilidad y validez para cerca del 31% de ellos, (4) el perfilador de estilos de aprendizaje de Jackson (LSP) sigue siendo un inventario sin evidencia en todos criterios de fiabilidad y validez.

Discusión y conclusiones. A pesar de la popularidad y aceptación de los estilos de aprendizaje, la recomendación inicial hecha hace 14 años por Coffield, Moseley, Hall, y Ecclestone, de no basar las intervenciones pedagógicas tomando como referencia únicamente a alguno de los instrumentos de estilos de aprendizaje sigue siendo válida.

Palabras Clave: Estilos de aprendizaje, inventarios, fiabilidad test-retest, validez convergente, validez factorial, consistencia interna, revisión sistemática.

Abstract

Introduction. One of the many notions that have been introduced in education in the last 50 years is that of learning styles. This idea is very popular and enjoys good acceptance. However, from a scientific perspective this notion is very controversial. In this sense, Coffield, Moseley, Hall, and Ecclestone, made the first systematic review on the subject, and showed that the psychometric instruments of learning styles do not have adequate levels of reliability and validity. However, the results of their review have not been updated since 2004. This contribution presents a systematic review of these aspects for the 2005-2010 period, with the aim of updating the information on the subject.

Method. Fifty-eight documents were examined, which contained enough information about the reliability and validity of the following inventories: Kolb's LSI, Felder's ILS, Sternberg's TSI, the VAK-VAKT-VARK sensory inventories, the inventories based on the model proposed by Entwistle (ASSIST, LASSI and RASI), Biggs' SPQ, Honey and Mumford's LSQ and Gregorc's Style Delineator (GSD).

Results. The analysis of the psychometric properties indicates that: (1) more than half of the inventories surveyed lack internal consistency and predictive validity, (2) around 31% have no test-retest reliability, (3) there is is no evidence on a at least one reliability and validity criteria for close to 31 % of them (4) Jackson's Learning Styles Profiler (LSP) remains an inventory without evidence on all reliability and validity criteria measures.

Discussion and Conclusion. Despite the popularity and acceptance of learning styles, the initial recommendation made 14 years ago by Coffield, Moseley, Hall, and Ecclestone, of not basing pedagogical interventions solely on any of the learning styles instruments is still valid.

Keywords: Learning styles, instruments, test-retest reliability, convergent validity, factorial validity, internal consistency, systematic review.

Introduction

One of the many notions that have been introduced in education in the last 50 years is that of learning styles. This idea is very popular and enjoys good acceptance. In a review of scientific publications during 2008, Lilienfield, Lynn, Ruscio, & Beyerstein (2010) counted up 1,984 articles in peer-reviewed journals, 919 presentations at conferences, and 701 books or book chapters on learning styles. A subsequent analysis for the 2013-2015 period, of 220 articles that were indexed in the ERIC and PubMed research databases, found that more than 85% of the literature begins and ends with a positive view of learning styles (Newton, 2015).

However, as has been shown in several areas (Gottfredson, 2009, Nirenburg, McShane, & Beale, 2004; Rao & Andrade, 2011), extensive citing, as well as the popularity and acceptance of ideas, methods, constructs and instruments does not imply that they are scientifically valid and produce positive results. In this sense, Coffield, Moseley, Hall, & Ecclestone (2004) conducted the first systematic review about the scientific validity of learning styles. This analysis found that there is no evidence in the literature that supports the hypothesis that matching the teaching strategy with the styles of the pupils improves learning, that is, no Aptitude-Treatment Interactions (ATI) were observed. In addition, the same revision does not recommend using the learning styles inventories, because they lack reliability and validity.

Goals and Hypothesis

More recent analyzes (Cuevas, 2015, Pashler, McDaniel, Rohrer, & Bjork., 2008) have confirmed the initial results with respect to ATI. However, the information on reliability and validity has not been updated since 2004. Systematic reviews and meta-anlyses should be updated regularly to prevent outdated scientific conclusions influencing decision-making, as well as future interventions (Lakens, Hilgard, Staaks, 2016). Consequently, this research presents an effort to update the reliability and validity matrix of Coffield et al. (2004), for articles published between 2005 and 2010, in the same way that Cuevas (2015) updated the review by Pashler et al. (2008) on the matching of teaching strategies and learning styles. The research question is the following: have the deficiencies of reliability and validity in the inventories of learning styles been corrected?

The rest of this document is organized as follows. First, the rules for the interpretation of reliability and validity statistics are introduced. Second, the literature search and selection

method is described. Third, the results of the systematic review are shown for each inventory. Finally, these results are discussed, and some conclusions are made.

Method

Procedure

Google Scholar (GS) was used for the literature search. It can be argued that GS is not a bibliographic or scientific literature database; however, the results of recent studies (Haddaway, Collins, Coughlin, & Kirk, 2015; Harzing & Alakangas, 2016), indicate that GS is at the same level in terms of coverage, and recovery of relevant literature as Scopus, Web of Science and PubMed. The string used for the search was the following: "learning AND (style OR styles) AND (inventory OR inventories) AND (validity OR reliability)". The first 160 GS results were reviewed year by year. In total, 99 documents were downloaded for the analysis. Those manuscripts that did not contain information on reliability and validity, did not addresse learning styles inventories, did not present enough statistical information or instruments that had less than 3 studies, were excluded from the final examination. Figure 1 shows this process.

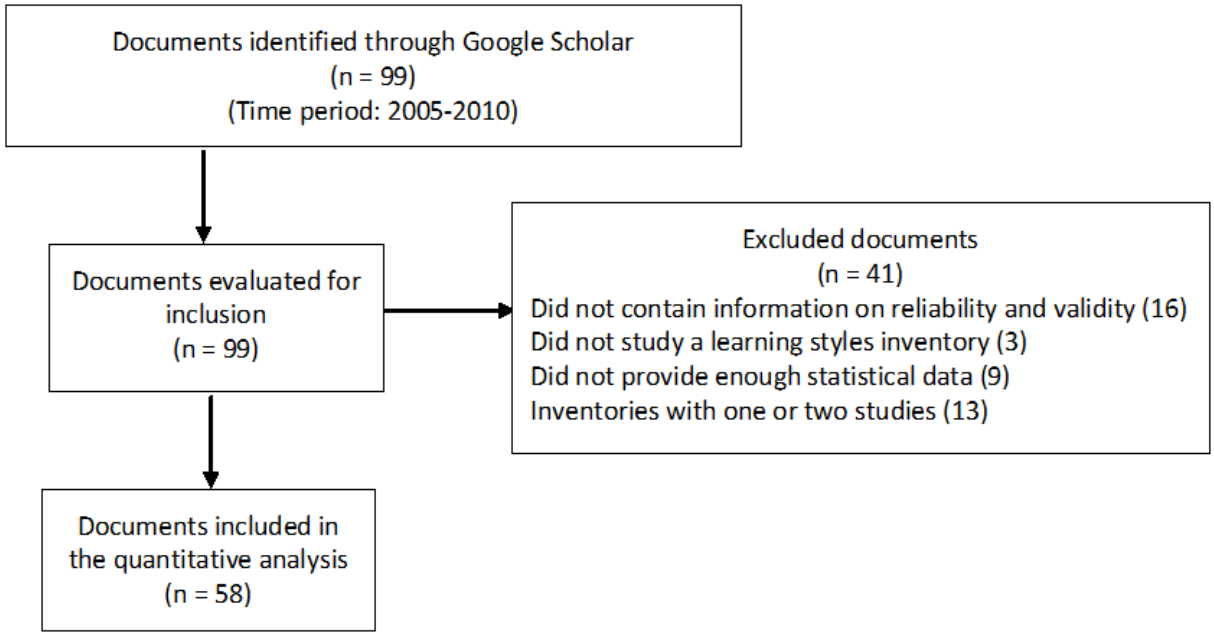

Figure 1. Literature search and selection flow diagram.

Data analysis

The threshold values that indicate acceptable internal consistency, test-retest reliability, as well as predictive and convergent validity, are shown in Table 1, according to the rec-

ommendations of Henson (2001), Lezak, Howieson, Bigler, & Tranel (2012), Mukaka (2012) and Carlson & Herdman (2012) respectively. With respect to the structural or factorial validity of the inventories, Table 2 presents the values that according to Iacobucci (2010) and Bagozzi & Yi (2012), goodness-of-fit indexes can take, to be considered good enough.

Table 1. *Threshold values for internal consistency, test-retest reliability, predictive validity and convergent validity*

Internal consitency	**Test-retest reliability**	**Predictive validity**	**Convergent validity**
$\alpha \geq .800$	$r \geq .700$	$r \geq .300$	$r \geq .700$

Table 2. *Threshold values for goodness-of-fit indexes of structural validity*

Relative chi-squared ($\chi 2/df$)	**Standardized Root Mean Residual (SRMR)**	**Root Mean Square Error of Approximation (RMSEA)**	**Non-Normed Fit Index (NNFI)**	**Comparative Fit Index (CFI)**
$\chi 2/df \leq 3$	SRMR≤.070	RMSEA≤.070	NNFI≥.920	CFI≥.930

For an inventory to have an acceptable consistency or reliability, all its scales must have coefficients values greater than or equal to those in Table 1. An instrument has adequate factorial validity if at least 2 fit indexes concur on this aspect (Bagozzi & Yi, 2012). Performing a Principal Component Analysis (PCA) to determine the structural validity of an inventory will be judged as lack of evidence. PCA is not a method for factorial analysis, and its multiple defects indicate that its use must be avoided (Costello & Osborne, 2005; Schmitt, 2011). Instead, an Exploratory Factorial Analysis (EFA) or a Confirmatory one (CFA) must be carried out. In case a study presents less than 2 fit indexes, it will be considered that there is no evidence of structural validity. The construct validity criteria can only be met, if the instrument shows both convergent and factorial validity. When there is only evidence of convergent or factorial validity, either for or against, construct validity will be marked as unresolved.

As part of the analysis, the partial omega squared statistic (ωp^2) will be calculated for all the F-tests (Lakens, 2013) and transformed into the correlation coefficient *r*. The results of the $\chi 2$ tests will also be converted into correlations using the standard formula (Rosenberg, 2010). When possible, correlations will be presented with their 95% confidence intervals.

Results

This section describes the information obtained from 8 instruments. They are presented in descending order from the inventory with more studies to the one with less studies.

Kolb's Learning Styles Inventory (LSI)

Kayes (2005) analyzed the internal consistency and structural validity of the LSI(*N*=221). Only 2 of the dimensions showed an adequate consistency: CE(α=.77), AC(α=.76), RO(α=.82) and AE(α=.82). Factorial validity was tested through a PCA.

Miller (2005) conducted two studies (*N*= 36 and *N*=34) to determine the predictive validity of the LSI. In the first study, it was determined that there was no effect of learning styles on the amount of material learned, $F(3,32)=.704$, $p=.557$, $\omega p^2=-.025$, $r=-.158$, 95% CI [-.463,.180], and in the result of the final evaluation, $F(3.32)=.821$, $p=.492$, $\omega p^2=-.015$, $r=-.123$, 95 % CI [-.443,.225]. None of the styles had a significant association with academic performance, with correlations ranging from $r=-.290$, $p>.05$, 95% CI [-.565,.043] to $r=.160$, $p>.05$, 95% CI [-.178,.464]. The second experiment indicated that there were only trivial and nonsignificant effects of styles in the amount of material learned, $F(3,30)=1.40$, $p=.262$, $\omega p^2=.034$, $r=.184$, 95% CI [- .164,.492] and the result of the final evaluation, $F(3.30)= .969$, $p=.420$, $\omega p^2=-.003$, $r=-.055$ 95% CI [-.386,.289]. In the latter study, the AC dimension had a significant negative correlation with the amount of material learned, $r=-.370$, $p<.05$, 95% CI [-.629, -. 036]. The rest of the styles only presented negligible and nonsignificant associations, that fluctuated between $r=-.240$, $p>.05$, 95% CI [-.535,.107] and $r=.290$, $p>.05$, 95% CI [-.552,.053].

Cooper Brathwaite (2006) measured the predictive validity of the LSI (*N*=76). The results indicate that the styles were not a significant predictor of the level of knowledge or the degree of cultural competence, showing correlations that varied between $r=.055$, $p= .640$, 95%, CI [-.173,.277] and $r=.141$, $p=.220$, 95% CI [-.087,.355].

Wang, Wang, Wang, & Huang (2006) tested the predictive validity of the LSI(*N*=455). The analysis points out that styles had a trivial but significant association with academic performance, $F(3,442)=6.81$, $p=.0002$, $\omega p^2=.038$, $r=.195$, 95% CI [.105,.282] .

Chapman & Calhoun (2006) examined the structural validity, as well as the convergent validity (*N*=94), between the LSI and Rezler & Removic's Learning Preferences Inventory (LPI). Only the abstract scales of the instruments (i.e., LSI-AC and LPI-AB) showed a

positive but insufficient correlation between them, r=.296, p<.05, 95% CI [.099,.470]. A PCA was used to test the factorial validity of the instrument.

Demirbas & Demirkan (2007) applied the LSI to 3 different groups (N=111, N= 88 and N=74), to examine its internal consistency and predictive validity. No acceptable consistency was observed in any of the groups. The coefficients oscillated between α=.51 and α=.73. Although one of the objectives of the study was to test the predictive validity of styles in relation to academic performance, no type statistical test result was reported to analyze the styles-performance association.

Yukselturk & Bulut (2007) investigated the predictive validity of the LSI(N=80). No style was a significant predictor of academic performance. The correlations ranged from r=-.080, p>.05, 95% CI [-.295,.142] and r=.150, p>.05, 95% CI [-.072,.358].

Lu, Jia, Gong, & Clark (2007) applied the LSI to evaluate its predictive validity (N=40). No significant association was found between learning styles and academic performance, $\chi 2$(3,N=40)=2.707, p=.538, r=.260, 95% CI [-.056,.529].

Akkoyunlu & Yilmaz-Soylu (2008) studied the predictive validity of the LSI(N=34). The results indicate that there is no association between styles and academic performance, F(1,32)=.389, p=.537, ωp^2=-.018, r=-.134 95% CI [-.452,.214].

Sun, Lin, & Yu (2008) examined the predictive validity of Kolb's inventory(N=56). A nonsignificant trivial association between styles and academic performance was found, F(3,51)=1.532, p=.217, ωp^2=.028, r=.167 95% CI [-.100,.412].

An & Yoo (2008) investigated the predictive validity of the inventory(N=742). It was established that styles have a trivial but significant relationship with critical thinking, r= .219, p<.001 95% CI [.149,.286].

Metallidou & Platsidou (2008) analyzed the internal consistency and structural validity of the LSI(N=338). The dimensions did not show an adequate consistency: CE(α=.77), RO(α=.78), AC(α=.73), AE(α=.78). Factorial validity was scrutinized through a PCA.

Andreou, Andreou, & Vlachos (2008) applied the LSI to analyze its internal consistency and predictive validity(N=452). The dependent variables were the phonological, syntactic and semantic aspects of verbal fluency. An acceptable level of consistency was observed in all dimensions: CE(α=.83), RO(α=.89), AC(α=.86), AE(α=.82). The correlations

between styles and the phonological aspect fluctuated between r=-.021, p>.05 95% CI [-.113,.071] and r=.246, p<.01, 95% CI [.157,.331]. Two styles presented a significant but trivial association with the syntactic aspect. The Accommodative style (AE + CE), r=.286, p<.01, 95% CI [.199,.369], and the Convergent (AC + AE) one, r=.250, p<.01, 95% CI [.161,.335]. A multiple regression corroborated the previous results for the phonological, $F(4,447)$= 2.58, p=.037, ωp^2=.014, r=.118 95% CI [.026,.208], syntactic, $F(4,447)$=4.54, p= .001, ωp^2=.030, r=.173 95% CI [.082, .261], and semantic, $F(4,447)$= 8.04, p=0.000003, ωp^2=.059, r=.243, 95% CI [.154,.328] aspects.

Brittan-Powell, Legum, & Taylor (2008) investigated the predictive validity of the LSI(N=102). A non-significant trivial effect between styles and academic performance as detected, $F(3,100)$=1.44, p=.47, ωp^2=.013, r=.114 95% CI [-.082,.302].

Cagiltay (2008) evaluated the predictive validity of Kolb's inventory(N=285). The results indicate a significant but trivial correlation between styles and academic performance, $F(3,281)$=6.06, p=.01, ωp^2=.051, r=.226 95% CI [.113,.333].

Erdem (2009) analyzed the internal consistency of the instrument(N=103). All the dimensions showed an inadequate consistency: CE(α=.58), RO(α=.70), AC(α=.71), AE(α=.65).

Yilmaz-Soylu & Akkoyunlu (2009) scrutinized the predictive validity of the LSI(N=139). There was no relationship between styles and academic performance, $F(1,37)$=.41, p=.526, ωp^2=-.015, r=-.123, 95% CI [-.422,.200].

Platsidou & Metallidou (2009) researched the internal consistency and structural validity of the inventory(N=340). Only one of the dimensions reached an acceptable consistency: CE(α=.81), RO(α=.72), AC(α=.76) and AE(α=.76). Factorial validity was tested by means of a PCA.

Jamieson (2010) measured the convergent validity of Kolb's instrument(N=68), with respect to 2 similar inventories developed with the Likert(N=34) and semantic differential(N=34) procedures. The correlations between the LSI and the other two instruments did not reach the desirable minimum level and oscillated between r=.402, p<.05, 95% CI [.225,.553] and r=.360, p<.05, 95% CI [.178, .518].

Azevedo & Akdere (2010) applied the LSI to determine its predictive validity to an experimental group(N=106) and a control group(N=96). The associations for the experimental

group oscillated between *r*=-.109, *p*> .05, 95% CI [-.294,.083] and *r*=.098, p> .05, 95% CI [-.095,.283]. While those in the control group fluctuated from *r*=-.059, *p*> .05, 95% CI [-.256, .143] to *r*=.177, *p*>.05, 95% CI [-.024,.365].

Table 3 summarizes the results presented above.

Table 3. *Summary of the results on reliability and validity for the LSI*

Study	**Internal consistency**	**Test-retest reliability**	**Construct validity**	**Predictive validity**
1. Kayes (2005)	YES	NE	NE	NE
2. Miller (2005)	NE	NE	NE	NO
3. Cooper Brathwaite (2006)	NE	NE	NE	NO
4. Wang et al. (2006)	NE	NE	NE	NO
5. Chapman & Calhoun (2006)	NE	NE	UR	NE
6. Demirbas y Demirkan (2007)	NO	NE	NE	NE
7. Yukselturk & Bulut (2007)	NE	NE	NE	NO
8. Lu et al. (2007)	NE	NE	NE	NO
9. Akkoyunlu & Yilmaz-Soylu (2008)	NE	NE	NE	NO
10. Sun, Lin, & Yu (2008)	NE	NE	NE	NO
11. An & Yoo (2008)	NE	NE	NE	NO
12. Metallidou & Platsidou (2008)	NO	NE	NE	SE
13. Andreou, Andreou, & Vlachos (2008)	NE	NE	NE	NO
14. Brittan-Powell, Legum, & Taylor (2008)	NE	NE	NE	NO
15. Cagiltay (2008)	NE	NE	NE	NO
16. Erdem (2009)	NO	NE	NE	NE
17. Yilmaz-Soylu & Akkoyunlu (2009)	NE	NE	NE	NO
18. Platsidou & Metallidou (2009)	NO	NE	NE	NE
19. Jamieson (2010)	NE	NE	UR	NE
20. Azevedo & Akdere (2010)	NE	NE	NE	NO

Note. NE (No Evidence); UR (Unresolved).

Felder's Index of Learning Styles (ILS)

Cook (2005) examined the internal consistency, test-retest reliability, and convergent validity of the ILS(N=138) with respect to Pelley's Learning Style Type Indicator (LSTI). No adequate consistency was observed: Active-Reflexive(α=.61), Sensing-Intuitive(α=.78), Visual-Verbal(α=.70), Sequential-Global(α=.67). The test-retest reliability almost reached an acceptable level: Active-Reflective (r= .75), Sensing-Intuitive (r= .81), Visual-Verbal (r= .60), Sequential-Global (r=.81). For convergent validity, the compared dimensions were Active-Reflective(ILS) vs. Extravert-Introvert(LSTI) and, Sensing-Intuitive(ILS) vs. Sensing-Intuitive(LSTI). The minimum acceptable degree of convergent validity was not reached with correlations that fluctuated between r=.60 and r=.69 respectively.

Cook & Smith (2006) measured the internal consistency, test-retest reliability, as well as the convergent inventory validity of the ILS(N=89), with respect to Kolb's LSI, Riding's Cognitive Style Analyzer (CSA) and Pelley's LSTI. None of the dimensions showed sufficient internal consistency: Active-Reflective(α=.62), Sensing-Intuitive(α=.77), Visual-Verbal(α=.72), Sequential-Global(α=.65). Two of the dimensions reached a good level of reliability: Active-Reflexive(r=.81), Sensing-Intuitive(r=.86), Visual-Verbal(r=.70) and Sequential-Global(r=.68). Only the Active-Reflective and Sensing-Intuitive facets, of both the LSTI and the ILS, had an almost acceptable convergence with correlations ranging from r=.681 and r=.684 respectively. The rest of the convergence values between the ILS, the LSI and the CSA, were well below the threshold value ($r \leq .495$), reaching negative values in the case of the Sequential-Global facet between the ILS and the CSA (r=-.206).

Galvan (2006) administered the ILS to study its predictive validity(N=84). None of the styles was a significant predictor of academic performance, with correlations oscillating between r=-.059, p=.421, 95% CI [-.270,.157] and r=.053, p=.398, 95% CI [-.163,.264].

Strang (2008) investigated the internal consistency and factorial validity of the ILS(N=715). Two of the dimensions did not not reach an adequate level of consistency: Active-Reflective(α=.857), Sensing-Intuitive(α=.786), Visual-Verbal(α=.827), Sequential-Global (α=.798). A PCA was carried out to analyze the structural validity.

Self & Widmann (2009) analyzed the predictive validity of ILS(N=86). Only null and trivial correlations between styles and students' level of conceptual knowledge were observed,

with associations that fluctuated between *r*=-.177, *p*=.102, 95% CI [-.375,.036] and *r*=.224, *p*=.038, 95% CI [.013,.416].

McChlery & Visser (2009) applied the ILS to 2 different samples of participants (*N*=469 and *N*=266), to measure its internal consistency. No acceptable level of consistency was observed in the first group: Active-Reflective(*α*=.50), Sensint-Intuitive(*α*=.60), Visual-Verbal(*α*=.63), Sequential-Global(*α*=.46). Nor in the second one: Active-Reflective(*α*=.53), Sensint-Intuitive(*α*=.63), Visual-Verbal(*α*=.66), Sequential-Global(*α*=.46).

Choi, Lee, & Kang (2009) investigated the predictive validity of the ILS(*N*=59). An association between styles and performance in problem solving was not found in any of the dimensios: Active-Reflective (*F*(8,106)=1.72, *p*=.10, ωp^2=.043, *r*=.207, 95% CI [-.052 ,.440]), Sensing-Intuitive (*F*(8,106)=.68, *p*=.71, ωp^2=-.023, *r*=-.152 95% CI [-.393,.108]), Visual-Verbal (*F*(8,106)=.52, *p*=.82, ωp^2=-.035, *r*=-.187 95% CI [-.073,.423]), Sequential-Global (*F*(8,106)=1.01, *p*= .43, ωp^2=.001, *r*=.032 95% CI [-.226,.286]).

Ku & Shen (2009) scrutinized the internal consistency and factorial validity of the ILS(*N*=1034). The instrument did not show an adequate consistency: Active-Reflective(*α*=.48), Sensing-Intuitive(*α*=.53), Visual-Verbal(*α*=.52), Sequential-Global(*α*=.41). The CFA could not validate the strucutre of 4 bipolar factors proposed by Felder, $\chi 2(896, N=1034) = 4636.37$, *p*<.001, $\chi 2/df$=5.175, *CFI*=.505.

Platsidou & Metallidou (2009) examined the internal consistency and factorial validity of the inventory(*N*=340). Consistency did not reach an adequate level: Active-Reflective(*α*=.45), Sensing-Intuitive (*α*=.62), Visual-Verbal (*α*=.51) and Sequential-Global(α=.45). A PCA was used as a factorial analysis method.

Zaharias, Andreou, & Vosinakis (2010) tested the predictive validity of the ILS(*N*=196). The results indicate that there is a negative but negligible relationship between styles and academic performance, *F*(7,188)=.375, *p*=.916, ωp^2=-.023, *r*=-.152 95% CI [-.286,-.012].

Alaoutinen (2010) applied the ILS to verify its predictive validity(*N*=84). Only the Active-Refelctive dimension had a significant non-trivial association with performance: Active-Reflective(*r*=.320, *p*=.05, 95% CI [.113,.500]), Sensing-Intuitive(*r*=.160, *p*=.320, 95% CI [-.056,.362]), Visual-Verbal(*r*=-.120, *p*=.470, 95% CI [-.326,.097]), Sequential-Global(*r*=-.010, *p*=.940, 95% CI [-.224,.205]).

Hosford & Siders (2010) examined the internal consistency, test-retest reliability and structural validity of the ILS (N = 358). No acceptable internal consistency was observed in any dimension: Active-Reflective(α=.76), Sensing-Intuitive(α=.62), Visual-Verbal (α=.64), Sequential-Global (α=.62). The test-retest correlations were obtained for 2 and 4 years. Only the Sensing-Intuitive dimension reached the desired minimum level: Active-Reflective(r=.58,r=.61), Sensing-Intuitive(r=.77,r=.72), Visual-Verbal(r=.65,r=.55) and Sequential-Global(r=.57,r=.47)). Factorial validity was tested through a PCA.

Clarke, Lesh, Trocchio, & Wolman (2010) studied the internal consistency, as well as the convergent validity of the instrument(N=86), with respect to Sternberg's Thinking Styles Inventory (TSI). None of the facets showed an acceptable level of consistency: Active-Reflective(α=.51), Sensing-Intuitive(α=.79), Visual-Verbal(α=.75), Sequential-Global(α=.59). Only the Sensing-Intuitive and Sequential-Global dimensions had significant non-trivial associations with the TSI. The first with the Executive(r=.408, p<.001 95% CI [.215,.571]), Conservative (r=.414, p<.001 95% CI [.222,.575]) and Local(r=.444, p<.001, 95% CI [.251,.603]) facets of the TSI. The second with the Conservative(r=.374, p<.001, 95% CI [.174,.545]) and Local (r=.513, p<.001, 95% CI [.334,.656]) ones of the same inventory. However, none of these correlations reached the desirable minimum level. Table 4 summarizes the results presented previously.

Table 4. *Summary of reliability and validity results for Felder's ILS*

Study	**Internal consistency**	**Test-retest reliability**	**Construct validity**	**Predictive validity**
1. Cook (2005)	NO	NO	UR	NE
2. Cook y Smith (2006)	NO	NO	UR	NE
3. Galvan (2006)	NE	NE	NE	NO
4. Strang (2008)	NO	NE	NE	NE
5. Self y Widmann (2009)	NE	NE	NE	NO
6. McChlery y Visser (2009)	NO	NE	NE	NE
7. Choi, Lee, y Kang (2009)	NE	NE	NE	NO
8. Ku y Shen (2009)	NO	NE	UR	NE
9. Platsidou y Metallidou (2009	NO	NE	NE	NE
10. Zaharias, Andreou, y Vosinakis	NE	NE	NE	NO

(2010)				
11. Alaoutinen (2010)	NE	NE	NE	NO
12. Hosford y Siders, (2010)	NO	NO	NE	NE
13. Clarke et al. (2010)	NO	NE	UR	NE

Note. NE (No Evidence); UR (Unresolved).

Sensory Styles Inventories (VAK, VAKT, VARK)

Unlike the 2 previous inventories, there is no single instrument for measuring sensory learning styles in the literature. Different authors propose different instruments. However, they coincide at least in the visual and auditory styles. Together they can be identified by the acronyms VAK, VAKT or VARK.

Isemonger & Sheppard (2007) administered Reid's Perceptual Learning Style Preference Questionnaire (PLSPQ), to examine its internal consistency and factorial validity(N=691). Only 2 out of 6 scales obtained an acceptable consistency: Visual(α=.37), Auditory(α=.39), Tactile(α=.67), Kinesthetic(α=.76), Group(α=.83), Individual(α=.84). The CFA did not confirm the 6-factor structure proposed by Reid, *SRMR*=.09, *RMSEA*=.09, *CFI*=.72.

Çirkinoglu & Demirci (2007) evaluated the predictive validity of Barsh's VAK inventory(N=573). The results indicate that that none of the styles had predictive validity with respect to academic performance: Visual(r=.037, p>.05, 95% CI [-.045,.119]), Auditory(r=.043, p>.05, 95% CI [-.039,.124]), Kinesthetic(r=.076, p> .05, 95% CI [-.006,.157]). The ANOVA test confirmed these results, $F(3,569)= 1.24$, p=.296, ωp^2=.001, r=.032 95% CI [-.050,.114].

Slack & Norwich (2007) investigated the internal consistency and test-retest reliability of Smith's VAK inventory(N=51). No scale reached an adequate degree of consistency: Visual (α=.63), Auditory (α=.75), Kinesthetic (α=.56). Only one of the dimensions did not have an acceptable level of reliability: Visual (r=.90), Auditory (r=.96), Kinesthetic (r=.75). However, it is important to underline that this last criterion was evaluated with a smaller sample (N=25).

Peters, Jones, & Peters (2008) applied Reid's PLSPQ, to analyze its internal consistency, as well as its predictive and factorial validity(N=338). Only 2 out of 6 scales obtained an acceptable consistency: Visual(α=.68), Auditory(α=.53), Tactile(α=.67), Kinesthetic(α=.72), Group(α=.91), Individual(α=.90). Correlations show that styles have little or no impact on academic performance: Visual(r=.102, p>.05, 95% CI [-.005,.206]), Auditory(r=.048, p>.05, 95% CI [-.059, .154]), Tactile(r=.025, p>.05, 95% CI [-.082,.131]), Kinesthetic(r=.075, p>

.05, 95% CI [-.032,.180]), Group(r=-.178, p<.01, 95% CI [-.279, -.073]), Individual (r=.247, p>.001 95% CI [.144,.345]). Structural validity was examined by means of a PCA.

Yildirim, Acar, Bull, & Sevinc (2008) measured the internal consistency and predictive validity of Reid's PLSPQ(N=746). Only 2 out of 6 scales obtained an acceptable consistency: Visual(α=.72), Auditory(α=.63), Tactile(α=.67), Kinesthetic(α=.64), Group(α=.80), Individual(α=.80). There were no interactions between styles and academic performance (see table 5).

Table 5. *Results of the F-test and correlations learning style-academic performance*

Dimension	F-test	Correlation
Visual	F(1,1443)=.179, p>.05, ω_p^2=-.00057	r=-.024 95% CI [-.096,.048]
Auditory	F(1,1443)=.053, p>.05, ω_p^2=-.00066	r=-.026 95% CI [-.098,.046]
Tactile	F(1,1443)=.039, p>.05, ω_p^2=-.00067	r=-.026 95% CI [-.098,.046]
Kinesthetic	F(1,1443)=.012, p>.05, ω_p^2=-.00068	r=-.026 95% CI [-.098,.046]
Group	F(1,1443)=.046, p>.05, ω_p^2=-.00066	r=-.026 95% CI [-.098,.046]
Individual	F(1,1443)=.087, p>.05, ω_p^2=-.00063	r=-.025 95% CI [-.097,.047]

Dobson (2009) tested the predictive validity of Fleming's VARK inventory(N=901). A trivial but significant relationship between styles and academic performance was observed, ηp^2=.05, r=.224, 95% CI [.161,.285].

Leite, Svinicki, & Shi (2010) examined the factorial validity of Fleming's VARK instrument(N=14,211). Three different methods were applied to carry out the CFA: CTCM, CTCU and CT-C(M-1). Both the CTCM (*RMSEA*=.047, *SRMR*=.064) and the CTCU (*RMSEA*=.034, *SRMR*=.049), verified the 4-factor structure proposed by Fleming. However, the third one could not, although the value of the *SRMR* was very close to the minimum required (*RMSEA*=.053, *SRMR*=.074). Table 6 summarizes the results presented previously.

Table 6. *Summary of reliability and validity results of sensory styles inventories*

Study	Internal consistency	Test-retest reliability	Construct validity	Predictive validity
1. Isemonger & Sheppard (2007)	NO	NE	UR	NE
2. Çirkinoglu & Demirci (2007)	NE	NE	NE	NO
3. Slack & Norwich (2007)	NO	NO	NE	NE
4. Peters, Jones, & Peters (2008)	NO	NE	NE	NO

5. Yildirim et al. (2008)	NO	NE	NE	NO
6. Dobson (2009)	NE	NE	NE	NO
7. Leite, Svinicki, & Shi, Y. (2010)	NE	NE	UR	NE

NotE. NE (No Evidence); UR (Unresolved).

Inventories for Learning and Study Skills and Strategies (ASSIST, LASSI, RASI)

Just as with the sensory styles inventories, there is no single instrument for learning and study skills and strategies. However, they are all based on Entwistle's approach (Coffield et al., 2004).

Cano (2006) administered the LASSI (Learning and Study Strategies Inventory) to 2 different samples (N=527 and N=429), and analized its internal consistency and factorial validity. Only 2 out of 10 dimensions showed a suitable level of consistency: Attitute(α=.61,α=.53), Movitation(α=.72,α=.68), Time Management(α=.84,α=.82), Anxiety(α=.78,α=.76), Concentration(α=.80,α=.79), Information Processing(α=.74,α=.73), Selecting Main Ideas(α=.66,α=.68), Study Aids(α=.62,α=.60), Self-testing(α=.63,α=.66), Test Strategies (α=.72, α=.73). Although the authors did not specify witch sample they used, the CFA could not corroborate the 3-factor structure proposed by the creators of the inventory, $\chi 2$(26)=96.67, p<.01; $\chi 2/df$=3.718; CFI=.95. However, a significant non-trivial relationship between styles and academic performance could be determined for the first, r=.326, p<.001, 95% CI [.247,.400], and second, r=.427, p<.001, 95% CI [.346,.501], group of participants.

Geertshuis & Fazey (2006) administered the RASI (Revised Approaches to Study Inventory), to measure its internal consistency and structural validity(N=389). None of the dimensions achieved adequate alpha levels: Deep(α=.70), Surface(α=.73), Strategic(α=.76). The latent structure of the instrument was determined by a PCA.

Diseth, Pallesen, Hovland, & Larsen (2006) applied the ASSIST (Approaches and Study Skills Inventory for Students), to examine its predictive and factorial validity(N=486). All dimensions had insignificant associations with academic performance: Deep(r=.190, p<.01, 95% CI [.103,.274]), Surface(r=-.180, p<.01, 95% IC [-.265,-.093]), Strategic(r=.240, p<.01, 95% CI [.154,.322]). The CFA verified the 3-factor structure proposed by the developers of the instrument, $\chi 2$(45,N=486)=120.86, p<.01, $\chi 2/df$=2,689; CFI=.96; $RMSEA$=.06.

Diseth (2007) investigated the internal consistency, predictive and factorial validity of the ASSIST inventory(*N*=206). Almost all dimensions did not have an acceptable consistency: Deep(*α*=.80), Surface(*α*=.74) and Strategic(*α*=.82). Only the Strategic style had a non-trivial positive association with academic performance: Deep(*r*=.250, *p*<.01, 95% CI [.117,.374]), Surface(*r*=-.450, *p*<. 01, 95% CI [-.553,-.334]), Strategic(*r*=.320, *p*<.01, 95% CI [.192,.438]). However, by controlling the effect of the Surface dimension, the contribution of the Strategic facet to performance was substantially reduced, *r*=.140, *p*<.01, 95% CI [.003, .272]. The CFA confirmed the 3-factor structure of the instrument, $\chi 2(45,N=206)=83.12$, $p<.05$; $\chi 2/df=1.85$, *CFI*=.95, *RMSEA*=.06.

Speth, Namuth, & Lee (2007) applied the ASSIST(*N*=446), to analyze its internal consistency and structural validity. No dimension had an acceptable consistency: Deep(*α*=.65), Surface(*α*=.70) and Strategic(*α*=.75). A PCA was used to test the structural validity.

Valadas, Gonçalves, and Faísca (2010) administered the ASSIST(*N*= 566), to evaluate its internal consistency. Two of the three dimensions presented an acceptable consistency: Deep(*α*=.81), Surface(*α*=.86) and Strategic(*α*=.79). Table 7 summarizes the results presented above.

Table 7. *Summary of reliability and validity results of the learning and study skills and strategies inventories*

Study	**Internal consistency**	**Test-retest reliability**	**Construct validity**	**Predictive validity**
1. Cano (2006)	NO	NE	UR	YES
2. Geertshuis & Fazey (2006)	NO	NE	NE	NE
3. Diseth et al. (2006)	NE	NE	UR	NO
4. Diseth, (2007)	NO	NE	NE	NO
5. Speth, Namuth, & Lee (2007)	NO	NE	NE	NE
6. Valadas, Gonçalves, & Faísca (2010)	NO	NE	NE	NE

NotE. NE (No Evidence); UR (Unresolved).

Sternberg's Thinking Styles Inventory (TSI)

Liu, Magjuka, & Lee (2008) measured the predictive validity of the TSI(*N*=208). The results indicate that there is no relationship between styles and academic performance, $F(2,206)=.447$, $p=.640$, $\omega p^2=-.005$, $r=-.071$, 95% CI [-.205,.066].

Black & McCoach (2008) analyzed the internal consistency and factorial validity of the TSI(*N*=798). Of the 12 subscales that the instrument has, 6 of them did not reach an adequate degree of consistency: Legislative(α=.81), Excecutive(α=.80), Judicial(α=.73), Monarchic(α=.64), Hierarchic(α=.80), Anarchic(α=.67), Local(α=.67), Global(α=.71), Internal (α=.79), External(α=.82) Liberal(α=.84), Conservative(α=.81). The CFA could not validate the 2-factor model proposed by the creators of the instrument, $\chi 2(26,N=798)=653.29$, $p<.001$, $\chi 2/df=25.12$, *CFI*=.779, *RMSEA*=.175.

Fan, Zhang, & Watkins (2010) administered the TSI(*N*=72), to examine its predictive validity. Eleven different tests of academic performance were applied. None of the dimensions was a predictor of performance, showing correlations below the minimum required (see Table 8). However, when controlling for personality and motivation aspects, the Anarchic, Executive, Judicial and Monarchic styles predicted 6 of the 11 performance tests, with correlations that fluctuated between $r=.316$, $p<.05$, 95% CI [.091, 510] and $r=.566$, $p<.001$, 95% CI [.385, .705].

Tabla 8. *Correlation intervals between thinking styles and academic performance*

Style	**Lowest correlation**	**Highest correlation**
Legislative	*r*=-.260 95% CI [-.464,-.030]	*r*=.020 95% CI [-.213,.251]
Executive	*r*=-.340 95% CI [-.530,-.118]	*r*=.010 95% CI [-.222,.241]
Judicial	*r*=-.310 95% CI [-.505,-.084]	*r*=.270 95% CI [.041,.472]
Monarchic	*r*=-.410 95% CI [-.586,-.197]	*r*=.030 95% CI [-.203,.260]
Hierarchic	*r*=-.250 95% CI [-.455,-.019]	*r*=.080 95% CI [-.155,.306]
Anarchic	*r*=-.240 95% CI [-.447,-.009]	*r*=.310 95% CI [.084, .505]
Local	*r*=-.190 95% CI [-.404,.044]	*r*=.170 95% CI [-.064,.386]
Global	*r*=-.210 95% CI [-.421,.023]	*r*=.210 95% CI [-.023,.421]
Internal	*r*=-.250 95% CI [-.455,-.019]	*r*=.280 95% CI [.052,

		.480]
External	*r*=-.290 95% CI [-.489,-.063]	*r*=.200 95% CI [-.033,.413]
Liberal	*r*=-.180 95% CI [-.395, .054]	*r*=.260 95% CI [.030, .464]
Conservative	*r*=-.300 95% CI [-.497,-.073]	*r*=.080 95% CI [-.155,.306]
Oligarchic	*r*=-.310 95% CI [-.505,-.084]	*r*=.130 95% CI [-.105,.351]

Clarke, Lesh, Trocchio, & Wolman (2010) applied the TSI (*N*=95) and analyzed their internal consistency, as well as the convergent validity of the instrument, with respect to Felder's ILS. Although Cronbach coefficients are not reported for all scales, none of the ones presented in the report, had an acceptable consistency: Judicial(α=.57), Liberal(α=.57), Executive(α=.79), Local(α=.79), Global(α=.63), Legislative(α=.64), Conservative(α=.69). Only 3 of the TSI styles (i.e., Executive, Conservative and Local), had non-trivial, positive significant correlations, with the ILS dimensions. However, the associations did not reach the desirable minimum value, ranging between *r*= .408, *p*<.001, 95% CI [.215,.571] and *r*=.444, *p*<.001, 95% CI [.251,.603]. Table 9 summarizes the results presented previously.

Tabla 9. *Summary of reliability and validity results for the TSI*

Study	**Internal consistency**	**Test-retest reliablity**	**Construct validity**	**Predictive validity**
1. Liu, Magjuka, & Lee (2008)	NE	NE	NE	NO
2. Black y McCoach (2008)	NO	NE	UR	NE
3. Fan, Zhang, & Watkins (2010)	NE	NE	NE	YES
4. Clarke et al. (2010)	NO	NE	NO	NO

NotE. NE (No Evidence); UR (Unresolved).

Biggs' Study Process Questionnaire (SPQ)

Furnham, Christopher, Garwood, & Martin (2007) applied the SPQ to examine its predictive validity(*N*=430). Only null or trivial associations were found with the student's general culture level, which fluctuated between *r*=-.270, 95% CI [-.355,-.180] and *r*=.150 95% CI [.056,.241]. The results of the logistic regression confirm these results,

$F(3,373)=9.57$, $\omega p^2=.064$, $r=.252$, 95% CI [.161,.338]. However, when the facets of motive and strategy of these styles were taken into account, the analysis showed a nontrivial positive relationship with performance, $F(6,367)=7.38$, $\omega p^2=.093$, $r=.305$, 95% CI [.217,.388].

Chamorro-Premuzic & Furnham (2008) investigated the internal consistency and predictive validity of the SPQ(N=158). None of the three dimensions had an adequate consistency: Surface(α=.66), Deep(α=.68), Achieving(α=.70). Only the Deep style had a low positive and significant association, with IQ, $r=.320$, $p<.01$, 95% CI [.172,.453], and the result of the final exam, $r=.330$, $p<.01$, 95% CI [.183,.462].

Chamorro-Premuzic & Furnham (2009) administered the SPQ to examine its internal consistency(N=852). None of the three dimensions had an acceptable consistency: Surface-Motive(α=.65), Deep-Motive(α=.66), Achieving-Motive(α=.68), Surface-Strategy(α=.65), Deep-Strategy(α=.75), Achieving-Strategy(α=.76).

Furnham, Monsen, & Ahmetoglu (2009) studied the predictive validity of SQP(N=212). The styles had no relation with the level of general culture of the students with correlations that fluctuated between $r=-.110$, 95% CI [-.241, .025] and $r= .130$, CI [-.005, .260]. The same results were observed with respect to the subjects of language and mathematics with associations that ranged from, $r= -.020$, 95% CI [-.115, .154] to $r=.180$, 95% CI [.046,.307]. Table 10 summarizes the results presented above.

Table 10. *Summary of reliability and validity results for the SPQ*

Study	Internal consistency	Test-restest reliability	Construct validity	Predictive validity
1. Furnham et al. (2007)	NE	NE	NE	YES
2. Chamorro-Premuzic & Furnham (2008)	NO	NE	NE	NO
3. Chamorro-Premuzic & Furnham (2009)	NO	NE	NE	NE
4. Furnham, Monsen, & Ahmetoglu (2009)	NE	NE	NE	NO

Note. NE (No Evidence); UR (Unresolved).

Honey & Mumford's Learning Styles Questionnaire (LSQ)

Rassool & Rawaf (2008) tested the predictive validity of the LSQ(*N*=110). The results indicate a null relationship between styles and students' knowledge acquisition, $F(2,106)=2,645$, $p= .076$, $\omega p^2=.029$, $r=.170$, 95% CI [-.018,.346].

Kappe, Boekholt, Den Rooyen, & Van der Flier (2009) examined the test-retest reliability and predictive validity of the LSQ(*N*=99). Only one of the dimensions presented an acceptable reliability: Activists($r=.70$), Reflectors($r=.63$), Theorists($r=.50$), Pragmatists($r=.46$). Four different measures were applied for academic performance. All associations between styles and performance were null with correlations that fluctuated between $r=-.070$, $p>.05$, 95% CI [-.264,.129] and $r=.150$, $p> .05$, 95% CI [-.049,.337].

Jackson, Hobman, Jimmieson, & Martin. (2009) analyzed the predictive validity of the LSQ(*N*=137). Styles only had trivial or null correlations with work performance, presenting associations that ranged between $r=-.020$, $p>.05$, 95% CI [-.187,.148] and $r=.140$, $p>. 05$, 95% CI [-.028,.301]. A multiple regression confirmed these results, $R^2=.030$, $r=.173$, 95% CI [.005,.331].

Sharif, Gifford, Morris, & Barber (2010) administered the LSQ(*N*=256), to investigate its predictive validity. The associations between styles and performance were null or trivial, varying between $r=-.200$, $p<.01$, 95% CI [-.315,-.079] and $r=.200$, $p<.01$, 95% CI [.079, .315]. Table 11 summarizes the results presented previously.

Tabla 11. *Summary of reliability and validity results for the LSQ*

Study	Internal consistency	Test-retest reliability	Construct validity	Predictive validity
1. Rassool & Rawaf (2008)	NE	NE	NE	NO
2. Kappe et al. (2009)	NE	NO	NE	NO
3. Jackson et al. (2009)	NE	NE	NE	NO
4. Sharif et al. (2010)	NE	NE	NE	NO

Nota. SE (No Evidence); UR (Unresolved).

Gregorc's Learning Style Delineator (GSD)

Miller (2005) applied the GSD to 2 different groups of participants (*N*=36 and *N*=34) and assessed its predictive validity, as well as its convergent validity with respect to Kolb's LSI. In the first group, only the Concrete-Random (CR) dimension had a nontrivial positive association with the final evaluation, *r*=.340, *p*<.05, 95% CI [.013,.601]. The correlations between the GSD and the LSI for this same group were null, fluctuating between r=-.300, p>.05, 95% CI [-.572,.032] and *r*=.320, *p*>.05, 95 % CI [-.010,.587]. In the second group, again the CR dimension had a positive relationship both with the amount of material learned, *r*=.420, *p*<.05, 95% CI [.095,.664] and with the final evaluation, *r*=.390, *p*<.05, 95% CI [.060,.643]. Only the Concrete-Sequential (CS) and CR dimensions of the GSD had positive relationships with the CE and RO ones of the LSI. The correlations between them were, *r*=.370, *p*<.05, 95% CI [.036,.629].

Ruhnau (2006) administered the GSD using 3 different samples (*N*=14, *N*=19 and *N*=9) to test its predictive validity. All the academic performance-style associations were null, both for the first, $F(3,10) = 2.280$, $p=.142$, $\omega p^2= .215$, $r=.464$, 95% CI [-.088,.798], second, $F(3.15)=.499$, $p=.689$, $\omega p^2= -.086$, $r=-.293$ 95% CI [-.659,.186] and third, $F(3.15)=.499$, $p=.689$, $\omega p^2= -.086$, $r=-.293$, 95% CI [-.659, .186], groups.

Reio & Wiswell (2006) examined the internal consistency and structural validity of the GSD(*N*=467). The instrument did not show an acceptable consistency: CS(α=.51), CR(α=.64), AR(α=.23), AS(α=.66). The CFA could not confirm the 4-factor structure proposed by Gregorc, $\chi 2(51,N=467)=407$, $\chi 2/df=7.99$, *CFI*=.51, *RMSEA*=.149. A bipolar 2-factor structure was additionally tested and could not be verified either, $\chi 2(49,N=467)=538$, $\chi 2/df=10.98$, *CFI*=.48, *RMSEA*=.191. Table 12 summarizes the results presented above.

Tabla 12. *Summary of reliability and validity results for the GSD*

Estudio	Consistencia interna	Fiabilidad test-retest	Validez de constructo	Validez predictiva
1. Miller (2005)	NE	NE	UR	NO
2. Ruhnau (2006)	NE	NE	NE	NO
3. Reio & Wiswell (2006)	NO	NE	UR	NO

Note. NE (No Evidence); UR (Unresolved).

Discussion and Conclusions

Three main conclusions can be made, based on the evidence presented previously. The first is that research, on the reliability and validity of learning styles inventories, is very dispersed. Most studies address at most, 2 of the 5 criteria needed to determine if an instrument is robust (i.e., internal consistency, test-retest reliability, predictive validity, convergent validity, and factorial validity). There is no systematic approach, on the part of researchers, to evaluate these aspects. In addition, it was found that around 55% of the investigations that sought to determine the factorial validity of some inventory, used a PCA for this purpose.

The second is that the 2 most studied crtieria of the inventories analyzed in this article, are predictive validity and internal consistency. In this sense, around 57% of the results analyzed indicate that none of the instruments has predictive validity. While about 43% point out that the instruments lack internal consistency.

The third is that the 2 least studied reliability and validity criteria, were test-retest reliability and construct validity. Ninety-one percent (91%) of the studies reviewed did not investigate the first aspect and 9% indicated that the instruments lack this type of reliability. About 77% of the researches did not analyze the second criterion and 21% could not solve it because they only examined either convergent or the factorial validity, but not both. For example, Jamieson (2010) studied the convergent validity of Kolb's LSI with 2 other similar inventories but did not address its factorial validity. In the same way, Reio & Wiswell (2006) analyzed the structural validity of Gregorc's GSD but did not study its convergent validity.

For comparison purposes, tables 13 and 14 show the reliability and validity matrix developed by Coffield et al. (2004), and the new updated and extended matrix, based on the results presented here. However, there is a small difference in the labels of the matrices. In the Coffield et al. matrix, the NE (No Evidence) label, includes both the absence of evidence, as well as limited or partial evidence, for or against. In this article, it was decided to add one more category, UR (Unresovled), which indicates that there is limited evidence, but that the matter has not been resolved yet. The labels with an asterisk are those that suffered modifications from one matrix to another, based on the analysis performed.

Table 13. *Coffield et al., reliability and validity matrix for learning style inventories*

Inventory	**Internal consistency**	**Test-retest reliability**	**Construct validity**	**Predictive validity**

1. Jackson	NE	NE	NE	NE
2. Riding	NO	NO	NO	NO
3. Sternberg	NO	NO	NO	NO
4. Dunn & Dunn	NO	NO	NO	YES
5. Gregorc	NO	NO	NO	YES*
6. Honey & Mumford	NO	YES	NO	NO
7. Kolb	NE*	YES	NO	NO
8. Entwistle	YES*	NE	YES	NO
9. Herrmann	NE	YES	YES	NE
10. Myer-Briggs	YES	YES	YES	YES
11. Apter	YES	YES	NE	YES
12. Vermunt	YES	YES	YES	NE
13. Allison y Hayes	YES	YES	YES	YES

Note. NE (No Evidence); UR (Unresolved).

Tabla 14. *New reliability and validity matrix for learning styles inventories*

Inventory	**Internal consistency**	**Test-retest reliability**	**Construc validity**	**Predictive validity**
1. Jackson	NE	NE	NE	NE
2. Riding	NO	NO	NO	NO
3. Sternberg	NO	NO	NO	NO
4. Dunn & Dunn	NO	NO	NO	YES
5. Gregorc	NO	NO	NO	NO*
6. Honey y Mumford	NO	YES	NO	NO
7. Kolb	NO*	YES	NO	NO
8. Entwistle	NO*	SE	SÍ	NO
9. Herrmann	NE	YES	YES	YES
10. Myer-Briggs	YES	YES	YES	YES
11. Apter	YES	YES	NE	YES
12. Vermunt	YES	YES	YES	NE
13. Allison y Hayes	YES	YES	YES	YES
14. Felder	NO	NO	UR	NO
15. VAK	NO	UR	UR	NO
16. Biggs	NO	NE	UR	NO

Note. NE (No Evidence); UR (Unresolved).

As can be seen, the reliability and validity of the inventories has changed very little. Of sixteen instruments, ten (63%) lack internal consistency and no evidence could be recovered for two (13%). Five instruments (31%) lack test-retest reliability, there is no evidence for four (25%) and in one (6%) it is unresolved. Six inventories (38%) have no construct validity, two lack evidence (13%) and 3 have not resolved the issue (19%). Nine inventories (56%) lack predictive validity and three (19%) have no evidence to support it. Only the inventory of Allison & Hayes, as well as that of Vermunt comply with (almost) all the criteria. Since only one article was retrieved for these inventories, they were not included in the analysis.

Overall, we can conclude that learning styles inventories are understudied, and their association with measures of performance is mostly null or trivial. Psychometrically, most of them limp. Depending on the instrument, they are robust in some aspect, such as internal consistency, and very weak in some other, like predictive validity. Consequently, the initial recommendation made 14 years ago by Coffield, Moseley, Hall, and Ecclestone (2004), of not basing pedagogical interventions solely on any of the learning styles instruments is still valid.

However, it is important to note that this research has six limitations. First, the search only included the 2005-2010 period. Second, only the first 160 GS results were reviewed. Third, no extra, ad-hoc searches were made for inventories that did not appear in the initial search. Fourth, for some studies, the evaluation of internal consistency and test-retest reliability was negative, because only one of its scales did not reach the minimum level required. Therefore, in these cases, the result should be interpreted with caution. Fifth, the different versions of each inventory were not considered. Although regardless of the inventory version, very similar results are consistently observed over time, for the period analyzed. Sixth, no analysis of the samples from each study, and the impact they have on the reliability and reproducibility of the results was carried out (Alvarez-Montero et al., 2018).

Finally, it can be argued that the criteria used in this research, to determine the reliability and validity of the inventories, are very stringent. In this sense, given that the promises to adopt the construct range from gains in academic performance to the development of respect for oneself and others (Dembo and Howard, 2007, Scott, 2010), it is useful to remember that extraordinary claims, require extraordinary evidence, based on the most rigorous standards (Voss, Helgen, and Jan-sa, 2014, Wagenmakers, Wetzels, Borsboom, and van der Maas, 2011). Otherwise, it is easy to confuse hopes and facts, and slide into pseudoscience (Sagan, 1997).

References

*Akkoyunlu, B., & Yilmaz-Soylu, M. (2008). A study of student's perceptions in a blended learning environment based on different learning styles. *Educational Technology & Society*, *11*(1), 183-193.

*Alaoutinen, S. (2010). Effects of learning style and student background on self-assessment and course performance. In *Proceedings of the 10th Koli Calling International Conference on Computing Education Research* (pp. 5-12). ACM. https://doi.org/10.1145/1930464.1930465

Alvarez-Montero, F.J.; Rocha-Ruiz, E.; Leyva-Cruz, M.G.; Moreno-Alcaraz, F.; Alvarez-Arredondo, A. (2018). The Instrumental Approach and Academic Achievement in Mathematics: A Systematic Review, Reliability Evaluation and Meta-analysis. *Proceedings of the International Conference on Education, Psychology and Learning (ICEPL-2018)*. Tokyo, Japan.

*An, G. J., & Yoo, M. S. (2008). Critical thinking and learning styles of nursing students at the baccalaureate nursing program in Korea. *Contemporary Nurse*, *29*(1), 100-109. https://doi.org/10.5172/conu.673.29.1.100

*Andreou, E., Andreou, G., & Vlachos, P. (2008). Learning styles and performance in second language tasks. *TESOL Quarterly, 42*(4), 665-674. https://doi.org/10.1002/j.1545-7249.2008.tb00156.x

*Azevedo, R. E., & Akdere, M. (2010). Measuring the effects of learning styles: Is a little knowledge dangerous for excellence in management education? *Journal of Leadership & Organizational Studies*, *17*(2), 192-200. https://doi.org/10.1177/1548051810366713

Bagozzi, R. P., & Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the academy of marketing science*, *40*(1), 8-34. https://doi.org/10.1007/s11747-011-0278-x

*Black, A. C., & McCoach, D. B. (2008). Validity study of the thinking styles inventory. *Journal for the Education of the Gifted*, *32*(2), 180-210. https://doi.org/10.4219/jeg-2008-849

*Brittan-Powell, C., Legum, H., & Taylor, E. (2008). The relationship between student learning style, selection of course delivery format, and academic performance. *International Journal of Instructional Technology and Distance Learning*, *5*(5), 41-46.

*Cagiltay, N. E. (2008). Using learning styles theory in engineering education. *European journal of engineering education*, *33*(4), 415-424. https://doi.org/10.1080/03043790802253541

*Cano, F. (2006). An in-depth analysis of the Learning and Study Strategies Inventory (LASSI). *Educational and Psychological Measurement*, *66*(6), 1023-1038. https://doi.org/10.1177/0013164406288167

Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, *15*(1), 17-32. https://doi.org/10.1177/1094428110392383

*Chamorro-Premuzic, T., & Furnham, A. (2008). Personality, intelligence and approaches to learning as predictors of academic performance. *Personality and individual differences*, *44*(7), 1596-1603. https://doi.org/10.1016/j.paid.2008.01.003

*Chamorro-Premuzic, T., & Furnham, A. (2009). Mainly Openness: The relationship between the Big Five personality traits and learning approaches. *Learning and Individual Differences*, *19*(4), 524-529. https://doi.org/10.1016/j.lindif.2009.06.004

*Chapman, D. M., & Calhoun, J. G. (2006). Validation of learning style measures: implications for medical education practice. *Medical education*, *40*(6), 576-583. https://doi.org/10.1111/j.1365-2929.2006.02476.x

*Choi, I., Lee, S. J., & Kang, J. (2009). Implementing a case-based e-learning environment in a lecture-oriented anaesthesiology class: Do learning styles matter in complex problem solving over time? *British Journal of Educational Technology*, *40*(5), 933-947. https://doi.org/10.1111/j.1467-8535.2008.00884.x

*Çirkinoglu, A. G., & Demirci, N. (2007). Correlation Between University Students' Kinematic Achievement and Learning Styles. In *Sixth International Conference of the Balkan Physical Union* (Vol. 899, pp. 485-486). https://doi.org/10.1063/1.2733248

*Clarke, T. A., Lesh, J. J., Trocchio, J. S., & Wolman, C. (2010). Thinking styles: teaching and learning styles in graduate education students. *Educational Psychology*, *30*(7), 837-848. https://doi.org/10.1080/01443410.2010.510794

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Learning styles and pedagogy in post-16 learning: A systematic and critical review*. London: Learning and Skills Research Centre.

*Cook, D. A. (2005). Reliability and validity of scores from the index of learning styles. *Academic medicine*, *80*(10), S97-S101.

*Cook, D. A., & Smith, A. J. (2006). Validity of index of learning styles scores: multitrait–multimethod comparison with three cognitive/learning style instruments. *Medical Education*, *40*(9), 900-907. https://doi.org/10.1111/j.1365-2929.2006.02542.x

*Cooper Brathwaite, A. (2006). Influence of nurse characteristics on the acquisition of cultural competence. *International Journal of Nursing Education Scholarship*, *3*(1). https://doi.org/10.2202/1548-923X.1173

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research & evaluation*, *10*(7), 1-9.

Cuevas, J. (2015). Is learning styles-based instruction effective? A comprehensive analysis of recent research on learning styles. *Theory and Research in Education*, *13*(3), 308-333. https://doi.org/10.1177/1477878515606621

Dembo, M. H., & Howard, K. (2007). Advice about the use of learning styles: A major myth in education. *Journal of college reading and learning*, *37*(2), 101-109. https://doi.org/10.1080/10790195.2007.10850200

*Demirbas, O. O., & Demirkan, H. (2007). Learning styles of design students and the relationship of academic performance and gender in design education. *Learning and instruction*, *17*(3), 345-359. https://doi.org/10.1016/j.learninstruc.2007.02.007

*Diseth, Å. (2007). Approaches to learning, course experience and examination grade among undergraduate psychology students: Testing of mediator effects and construct validity. *Studies in Higher Education*, *32*(3), 373-388. https://doi.org/10.1080/03075070701346949

*Diseth, Å., Pallesen, S., Hovland, A., & Larsen, S. (2006). Course experience approaches to learning and academic achievement. *Education+ Training*, *48*(2/3), 156-169. https://doi.org/10.1108/00400910610651782

*Dobson, J. L. (2009). Learning style preferences and course performance in an undergraduate physiology class. *Advances in Physiology Education*, *33*(4), 308-314. https://doi.org/10.1152/advan.00048.2009

*Erdem, M. (2009). Effects of learning style profile of team on quality of materials developed in collaborative learning processes. *Active Learning in Higher Education*, *10*(2), 154-171. https://doi.org/10.1177/1469787409104902

*Fan, W., Zhang, L. F., & Watkins, D. (2010). Incremental validity of thinking styles in predicting academic achievements: An experimental study in hypermedia learning envi-

ronments. *Educational Psychology*, *30*(5), 605-623. https://doi.org/10.1080/01443410.2010.496899

*Furnham, A., Christopher, A. N., Garwood, J., & Martin, G. N. (2007). Approaches to learning and the acquisition of general knowledge. *Personality and Individual Differences*, *43*(6), 1563-1571. https://doi.org/10.1016/j.paid.2007.04.013

*Furnham, A., Monsen, J., & Ahmetoglu, G. (2009). Typical intellectual engagement, Big Five personality traits, approaches to learning and cognitive ability predictors of academic performance. *British Journal of Educational Psychology*, *79*(4), 769-782. https://doi.org/10.1348/978185409X412147

*Galvan, M. T. (2006). Can students' learning style preferences explain their academic performance in marketing courses? *Proceedings of the Marketing Management Association*, 117-124.

*Geertshuis, S. A., & Fazey, J. A. (2006). Approaches to learning in the workplace. *Journal of Workplace Learning*, *18*(1), 55-65. https://doi.org/10.1108/13665620610641319

Gottfredson, L. S. (2009). Logical fallacies used to dismiss the evidence on intelligence testing. In R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing* (pp. 11-65). Washington, DC, US: American Psychological Association. http://doi.org/10.1037/11861-001

Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PloS one*, *10*(9), e0138237. https://doi.org/10.1371/journal.pone.0138237

Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*(2), 787-804. https://doi.org/10.1007/s11192-015-1798-9

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and evaluation in counseling and development*, *34*(3), 177.

*Hosford, C. C., & Siders, W. A. (2010). Felder-Soloman's index of learning styles: Internal consistency, temporal stability, and factor structure. *Teaching and learning in medicine*, *22*(4), 298-303. https://doi.org/10.1080/10401334.2010.512832

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*(1), 90-98. https://doi.org/10.1016/j.jcps.2009.09.003

*Isemonger, I., & Sheppard, C. (2007). A construct-related validity study on a Korean version of the perceptual learning styles preference questionnaire. *Educational and Psychological Measurement*, *67*(2), 357-368. https://doi.org/10.1177/0013164406292073

*Jackson, C. J., Hobman, E. V., Jimmieson, N. L., & Martin, R. (2009). Comparing different approach and avoidance models of learning and personality in the prediction of work, university, and leadership outcomes. *British Journal of Psychology*, *100*(2), 283-312. https://doi.org/10.1348/000712608X322900

*Jamieson, G. E. (2010). *Investigating the Kolb Learning Style Inventory's ipsative scores using semantic differential and Likert scaling* (Doctoral dissertation, Northern Michigan University).

*Kappe, F. R., Boekholt, L., Den Rooyen, C., & Van der Flier, H. (2009). A predictive validity study of the Learning Style Questionnaire (LSQ) using multiple, specific learning criteria. *Learning and Individual differences*, *19*(4), 464-467. https://doi.org/10.1016/j.lindif.2009.04.001

*Kayes, D. C. (2005). Internal validity and reliability of Kolb's learning style inventory version 3 (1999). *Journal of Business and Psychology*, *20*(2), 249-257. https://doi.org/10.1007/s10869-005-8262-4

*Ku, D. T., & Shen, C. Y. (2009). Reliability, validity, and investigation of the index of learning styles in a Chinese language version for late adolescents of Taiwanese. *Adolescence*, *44*(176), 827.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, *4*. ttps://doi.org/10.3389/fpsyg.2013.00863

Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC psychology*, *4*(1), 24. https://doi.org/10.1186/s40359-016-0126-3

*Leite, W. L., Svinicki, M., & Shi, Y. (2010). Attempted validation of the scores of the VARK: Learning styles inventory with multitrait–multimethod confirmatory factor analysis models. *Educational and psychological measurement*, *70*(2), 323-339. https://doi.org/10.1177/0013164409344507

Lezak, M. D., Howieson, D., Bigler, E.D., & Tranel, D. (2012). Neuropsychological assessment. 5th edition. Oxford University Press. *Oxford, New York, ISBN*, *10*, 9780195395525.

Lilienfield, S.O., Lynn, S.J., Ruscio, J., & Beyerstein B.L. (2010). *50 great myths of popular psychology—Shattering widespread misconceptions about human behavior.* West Sussex, UK: Wiley-Blackwell.

*Liu, X., Magjuka, R. J., & Lee, S. H. (2008). The effects of cognitive thinking styles, trust, conflict management on online students' learning and virtual team performance. *British Journal of Educational Technology*, *39*(5), 829-846. https://doi.org/10.1111/j.1467-8535.2007.00775.x

*Lu, H., Jia, L., Gong, S. H., & Clark, B. (2007). The relationship of Kolb learning styles, online learning behaviors and learning outcomes. *Journal of Educational Technology & Society*, *10*(4).

*McChlery, S., & Visser, S. (2009). A comparative analysis of the learning styles of accounting students in the United Kingdom and South Africa. *Research in Post-Compulsory Education*, *14*(3), 299-315. https://doi.org/10.1080/13596740903139404

*Metallidou, P., & Platsidou, M. (2008). Kolb's Learning Style Inventory-1985: Validity issues and relations with metacognitive knowledge about problem-solving strategies. *Learning and Individual Differences*, *18*(1), 114-119. https://doi.org/10.1016/j.lindif.2007.11.001

*Miller, L. M. (2005). Using learning styles to evaluate computer-based instruction. *Computers in human behavior*, *21*(2), 287-306. https://doi.org/10.1016/j.chb.2004.02.011

Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, *24*(3), 69-71.

Nirenburg, S., McShane, M., & Beale, S. (2004). The rationale for building resources expressly for NLP. In *Proc. of the 4th International Conference on Language Resources and Evaluation* (Vol. 218).

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological science in the public interest*, *9*(3), 105-119. https://doi.org/10.1111/j.1539-6053.2009.01038.x

*Peters, D., Jones, G., & Peters, J. (2008). Preferred 'learning styles' in students studying sports-related programmes in higher education in the United Kingdom. *Studies in Higher Education*, *33*(2), 155-166. https://doi.org/10.1080/03075070801916005

*Platsidou, M., & Metallidou, P. (2009). Validity and Reliability Issues of Two Learning Style Inventories in a Greek Sample: Kolb's Learning Style Inventory and Felder & Soloman's Index of Learning Styles. *International Journal of Teaching and Learning in Higher Education*, *20*(3), 324-335.

Rao, T. S., & Andrade, C. (2011). The MMR vaccine and autism: Sensation, refutation, retraction, and fraud. *Indian journal of psychiatry*, *53*(2), 95. https://doi.org/10.4103/0019-5545.82529

*Rassool, G. H., & Rawaf, S. (2008). The influence of learning styles preference of undergraduate nursing students on educational outcomes in substance use education. *Nurse Education in Practice*, *8*(5), 306-314. https://doi.org/10.1016/j.nepr.2008.02.001

*Reio Jr, T. G., & Wiswell, A. K. (2006). An examination of the factor structure and construct validity of the Gregorc Style Delineator. *Educational and Psychological Measurement*, *66*(3), 489-501. https://doi.org/10.1177/0013164405282459

Rosenberg, M. S. (2010). A generalized formula for converting chi-square tests to effect sizes for meta-analysis. *PloS one*, *5*(4), e10059. https://doi.org/10.1371/journal.pone.0010059

*Ruhnau, K. (2006). An Analysis of Learning Outcomes of Adult Students: Learning Styles. *Statistics [NCES]*, *2002*, 0-12.

Sagan, C. (1997). *The demon-haunted world: Science as a candle in the dark*. Random House Digital, Inc.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304-321. https://doi.org/10.1177/0734282911406653

Scott, C. (2010). The enduring appeal of 'learning styles'. *Australian Journal of Education*, *54*(1), 5-17. https://doi.org/10.1177/000494411005400102

*Self, B. P., & Widmann, J. (2009). Work in progress-learning styles and performance on the dynamics concept inventory. In *Frontiers in Education Conference, 2009. FIE'09. 39th IEEE* (pp. 1-2). IEEE. https://doi.org/10.1109/FIE.2009.5350822

*Sharif, S., Gifford, L. A., Morris, G. A., & Barber, J. (2010). The relationship between learning styles, attendance and academic performance of Pharmacy undergraduates. *Pharmacy Education*, *10*.

*Slack, N., & Norwich, B. (2007). Evaluating the reliability and validity of a learning styles inventory: A classroom-based study. *Educational Research*, *49*(1), 51-63. https://doi.org/10.1080/00131880701200765

*Speth, C. A., Namuth, D. M., & Lee, D. J. (2007). Using the ASSIST Short Form for Evaluating an Information Technology Application: Validity and Reliability Issues. *Informing Science*, *10*.

*Strang, K. D. (2008). Quantitative online student profiling to forecast academic outcome from learning styles using dendrogram decision models. *Multicultural Education & Technology Journal*, *2*(4), 215-242. https://doi.org/10.1108/17504970810911043

*Sun, K. T., Lin, Y. C., & Yu, C. J. (2008). A study on learning effect among different learning styles in a Web-based lab of science for elementary school students. *Computers & Education*, *50*(4), 1411-1422. https://doi.org/10.1016/j.compedu.2007.01.003

*Valadas, S. C. A., Gonçalves, F. R., & Faísca, L. M. (2010). Approaches to studying in higher education Portuguese students: a Portuguese version of the approaches and study skills inventory for students. *Higher Education*, *59*(3), 259-275. https://doi.org/10.1007/s10734-009-9246-5

Voss, R. S., Helgen, K. M., & Jansa, S. A. (2014). Extraordinary claims require extraordinary evidence: a comment on Cozzuol et al. (2013). *Journal of Mammalogy*, *95*(4), 893-898. https://doi.org/10.1644/14-MAMM-A-054

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & Van Der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). https://doi.org/10.1037/a0022790

*Wang, K. H., Wang, T. H., Wang, W. L., & Huang, S. C. (2006). Learning styles and formative assessment strategy: enhancing student achievement in Web-based learning. *Journal of Computer Assisted Learning*, *22*(3), 207-217. https://doi.org/10.1111/j.1365-2729.2006.00166.x

*Yildirim, O., Acar, A. C., Bull, S., & Sevinc, L. (2008). Relationships between teachers' perceived leadership style, students' learning style, and academic achievement: A study on high school students. *Educational Psychology*, *28*(1), 73-81. https://doi.org/10.1080/01443410701417945

*Yilmaz-Soylu, M., & Akkoyunlu, B. (2009). The effect of learning styles on achievement in different learning environments. *TOJET: The Turkish Online Journal of Educational Technology*, *8*(4).

*Yukselturk, E., & Bulut, S. (2007). Predictors for student success in an online course. *Journal of Educational Technology & Society*, *10*(2).

*Zaharias, P., Andreou, I., & Vosinakis, S. (2010). Educational virtual worlds, learning styles and learning effectiveness: An empirical investigation. In *Proceedings of the 7th Pan-Hellenic Conference with International Participation ICT in Education, Korinthos, Greece* (pp. 1-6).