إستخدام الأساليب الكمية فى التحقق من صدق محتوى الاختبار النفسى

د. **محسوب عبد القادر الضوى حسن** *

مقدمة

تعتبر دراسة الصدق فى البحوث النفسية أهم خطوة من خطوات إعداد الاختبار النفسى ، وتعتبر من أهم خصائصه الأساسية التى ينبغى أن تتوافر ، ولا يمكن الاستغناء عنها مهما كانت شروط الاختبار الجيد الأخرى كالثبات متوافر بدرجة كبيرة ، حيث يمثل الصدق مسألة جوهرية فى الاختبارات لها متطلباتها وشروط التحقق منها ، ولأن القياس النفسى قياس غير مباشر فإنه يتطلب التأكد من أن ما تقيسه الأدوات يمكن الثقة فيه والاعتماد عليه فى جمع البيانات .

وتستند قوة تصميم الدراسة البحثية بشكل كبير على درجة دقة قياس المتغيرات المستهدفة ، لذا يشير الصدق إلى أى مدى يمكن أن تقيس البنود المحددة فى أداة القياس المفهوم المستهدف بالقياس، ويضمن أن البنود تسمح بعمل استدلالات صادقة (Sangoseni, Hellman & Hill, 2013: 2)

ويعتمد البحث الامبيرقى على الفحص المنتظم لتجريدات مفاهيمية Conceptua Abstracts خلال استجابات قابلة للملاحظة والقياس. وتستخدم هذه العملية لتحديد وتفسير الظاهرة موضع الاهتمام . ويُعد صدق المحتوى Content Validity خطوة مهمة فى بناء أدوات قياس

* أستاذ مساعد علم النفس التربوى كلية التربية بقنا – جامعة جنوب الوادى

جديدة لأنه يمثل آلية أولية لربط المفاهيم المجردة مع مؤشرات قابلة للملاحظة والقياس (Wynd, Schmidt & Schaefer, 2003: 508) .

ومنذ نشر معايير بناء الاختبار Standards for Test Construction (1954/1999) التى أعدتها الجمعية الأمريكية لعلم النفس ورابطة البحث التربوى الأمريكية والمجلس الوطنى للقياس فى التربية كان هناك إصرار متزايد على الحاجة إلى توفير دليلاً على أن الاختبارات التى يتم تطويرها تمثل على نحو ملائم مجالات محتوى البنية المستهدفة بالقياس . وعلى وجه التحديد أن تعكس مختلف معايير بناء الاختبار الأهمية المتزايدة المعطاة لثلاثة مفاهيم متلازمة فى مجال بناء الاختبار وهى : التعريف Definition ، والتمثيل Representativeness ، والارتباط Relevance . وتُشكل هذه المفاهيم ، البنية التصورية الأساسية لتقويم صدق المحتوى (Delgado-Rico, Carretero-Dios & Ruch, 2012: 450)

كما حمل الأدب السيكولوجى تعريفاً أوسع لصدق المحتوى يشمل مظاهر من بناء الاختبار ، حيث وُصف بأنه يتعلق بأربعة عناصر لجودة الاختبار هى : تعريف المجال ، تمثيل المجال ، وارتباط المجال ، وملاءمة إجراءات بناء الاختبار ، وتوفر هذه العناصر إطار عمل لتقويم محتوى الاختبار (Sireci,1998b;Sireci& Faulkner-Bond, 2014: 100)

ويتطلب صدق المحتوى كلاً من صدق البند وصدق المعاينة ، ويتعلق صدق البند بما إذا كانت بنود الاختبار موائمة لقياس مجال المحتوى المرجو ، ويتعلق صدق المعاينة Sampling Validity بجودة معاينة الاختبار للمجال الكلى Domain Universe للمحتوى موضع القياس ، أى أن العينة (البنود الحالية للاختبار) تمثل مجتمع الأصل (كل البنود الممكنة) (محمود أحمد عمر وحصة عبد الرحمن فخرو

وتركى السبيعى وآمنة عبد الله تركى ، ٢٠١٠ : ١٩٢ ؛ جاى وميلز وإيراسيان ، ٢٠١٢ :٢٥٢) .

ومن الوجهة التاريخية ، استخدم صدق المحتوى فى بادىء الأمر مع الاختبارات التحصيلية حيث تكون الأسئلة ممثلة لجميع عناصر المحتوى أفضل تمثيل وبالتالى فإن الاختبار يكون ذا صدق ومحتوى عال . أما فى حالة الاختبارات السيكولوجية فإن الأمر يختلف بعض الشىء حيث يتطلب الأمر تحليل أعمق لمجال القياس ، وكذلك تحليل أعمق لمحتوى القدرة أو السمة لمعرفة مكوناتها من خلال تقديرات مجموعة من المحكمين المتخصصين ، كما أنه أكثر عرضة لأخطاء التقدير رغم الاعتماد عليه فى البحوث وذلك بسبب غياب المؤشرات الإحصائية لفحص الاتفاق بين المحكمين (أمين على محمد سليمان ، ٢٠١٠ : ٥٨٦ ؛ Janda, 1998: 80) .

ويوجد شبه إتفاق فى أدبيات مناهج البحث على أن صدق المحتوى مسألة إصدار حُكم بشكل كبير تشتمل على خطوتين متمايزتين الأولى : تتضمن جهد مُسبق من مُعد الاختبار لتعزيز صدق المحتوى من خلال عملية تصورية دقيقة Careful Conceptualization وتحليل محتوى قبل توليد البنود ، والثانية : تتضمن جهد لاحق لتقويم الارتباط بين محتوى الاختبار من خلال تقييم خبير (محكم) (Carmines & Zeller, 1979: 22; Popham, 1990: 98; Beck & Gable, 2001; Lynn, 1986: 382-383; Mastaglia, Toye & Kristjanson, 2003; Bayrak & Yurdugul, 2012: 264; Zamanzadeh, et al., 2014: 163) .

وبمرور السنوات بدأ الباحثون تحديد الحاجة إلى اختبار مطابقة البنود المشتقة كيفياً بمجالات المحتوى الخاصة بها ، وبذلت جهود لإيجاد المزيد من الطرق القابلة للتكميم لتحديد صدق المحتوى . وقد

أثمرت جهود المتخصصين فى القياس النفسى عن مجموعة من طرق تكمية صدق المحتوى منها : مؤشر صدق المحتوى Content Validity Index (CVI) والتى تُسمى طريقة نسبة الاتفاق Proportion Agreement Method ، ومعامل الاتفاق Coefficient of Agreement ، ومعامل كابا للإتفاق متعدد المقدرين Multirater Kappa Coefficient of Agreement ، ومؤشر الصدق العاملى Factorial Validity Index (FVI) ، وفترة الثقة غير المتماثلة للمتوسط Asymmetric Confidence Interval for the Mean ، ومعامل صدق المحتوى Content Validity Coefficient (VIk) ، ومتوسط نسبة التطابق Average of Congruency Percentage (ACP) ، ونسبة صدق المحتوى Content Validity Ratio (CVR) ، ومؤشر تطابق البند Index of Item Congruence ، وتحليل التجزىء الكامن Latent Partition Analysis (Lu, 1971; Lawshe, 1975; Popham, 1978; Rovinelli & Hambleton, 1977; Tinsley & Weiss, 1975; Waltz & Bausell, 1981; Aiken, 1985; James, Demaree & Wolf, 1984; Lynn, 1986; Cohen, 1960; Wynd et al., 2003: 509; Miller & Penfield, 2005) .

وقد قدم Fitzpatrick (1983: 3-10) تحليلاً وافياً لست طرق مستقلة لتقويم صدق المحتوى تركز أربع منها على بنود الاختبار وهى : وضوح تعريفات مجال المحتوى Clarity ، وارتباط محتوى الاختبار بمجال المحتوىRelevance ، وملاءمة (كفاية) المعاينة Sampling Adequacy لمحتوى الاختبار ، والجودة الفنية Technical Quality لبنود الاختبار . أما الآخران فيركزان على المستجيب على الاختبار – الأول هو ملاءمة المعاينة للاستجابات على الاختبار ، والثانى ارتباط الاستجابات على الاختبار ببنية سلوكية .

كما حدد Yaghmaie (2003: 27) أربعة محكات لقياس صدق المحتوى هى : الارتباط (انتماء بنود الاختبار لمجال المحتوى) ، ووضوح البنود ، والبساطة Simplicity ، والغموض Ambiguity . فى حين اقترح (Gholami, Pakdaman, Jafari & Virtanen, 2014: 197) أن يُقوم صدق المحتوى بواسطة لجنة من الخبراء فيما يتعلق بالانتماء ، والوضوح ، والبساطة ، وضرورة البنود Necessity of Items .

وقد اتجهت البحوث والدراسات الحديثة إلى تكمية صدق المحتوى باستخدام صيغة إحصائية أو أكثر ومنها على سبيل المثال : دراسة Treloar, Martens and McCarthy (2015) التى استخدمت نسبة صدق المحتوى CVR فى مقياس استراتيجيات الحماية السلوكية ، ودراسة van Schelven et al. (2015) التى استخدمت مؤشر صدق المحتوى CVI لتكمية محك وضوح الصياغة فى الصورة الألمانية من مسح اتجاه الممرضات نحو مرضى الرعاية الحرجة كبار السن ، ودراسة Nordin, Elf, McKee and Wijk (2015) التى استخدمت مؤشر صدق المحتوى فى الصورة السويدية من مسح تقييم بيئة الرعاية لمرضى الرعاية الحرجة ، ودراسة Vadlin, Aslund and Nilsson (2015) التى استخدمت مؤشر صدق المحتوى فى مسح فرز إدمان الألعاب لدى المراهقين ، ودراسة Farahani, Ghasemi, Nikpeyma, Fereidouni and Rassouli (2015) التى استخدمت مؤشر ونسبة صدق المحتوى معاً فى قائمة الأداء التدريسى الكلينيكى للمعلمين ، ودراسة Lotfi, Tehrani, Yaghmaie and Hajizadeh (2014) التى استخدمت مؤشر ونسبة صدق المحتوى معاً فى استبيان سلوكيات الحماية الأنثوية للنساء المتعرضات للإصابة بمرض الإيدز ، ودراسة Yuen et al. (2014) التى استخدمت مؤشر صدق المحتوى لتكمية محكا

الانتماء والوضوح فى مقياس محو الأمية الصحية لمقدمى الرعاية لمرضى السرطان ، ودراسة (2013) .Ebn Ahmady et al التى استخدمت مؤشر ونسبة صدق المحتوى معاً فى الصورة الفارسية من استبيان معرفة التقييم الشامل ، ودراسة (2013) Yaghmaie, Mohammadi and Majd التى استخدمت مؤشر صدق المحتوى فى استبيان جودة الحياة لدى الأزواج المحرومين من الإنجاب .

وتأتى الدراسة الحالية لإلقاء الضوء على خمس صيغ إحصائية مطروحة فى أدبيات القياس النفسى لتكمية صدق المحتوى نظرياً وإمبريقياً من خلال استخدامها لفحص محكات الوضوح والبساطة والارتباط والغموض لثلاثة اختبارات نفسية معدة لهذا الغرض .

مشكلة الدراسة

لكى تؤدى الاختبارات وظيفتها وتحقق الهدف الذى أعدت من أجله باعتبارها وسائل للقياس توافر عامل الصدق فهو عامل بالغ الأهمية فى اختيار وتطبيق أدوات القياس ، لذا من المهم أن يتأكد مُعد/مستخدم الاختبار من صدقه الذى يُعد أكثر خصائص الاختبار أهمية . وبدونه لا يكون لتفسيرات البيانات معنى أو يكون معناها غير مناسب .

ويُعد الصدق مهم فى جميع أشكال البحوث وكل أنواع الاختبارات والمقاييس ، ومن الأفضل النظر إليه من حيث الدرجة : صادق بدرجة كبيرة ، صادق بدرجة متوسطة ، وغير صادق بعامة . ويبدأ تقدير الصدق بفهم التفسير أو التفسيرات المستمدة من اختبارات أو أدوات منتقاة . ومن ثم يتطلب جمع عدد من الأدلة لتأييد التفسير المرجو. ومستخدم الاختبار هو الذى يتخذ القرار النهائى فيما يتعلق بصدق وفائدة اختبار أو مقياس معين ، وينبغى وصف الأساس الذى

استند إليه القرار فى قسم الإجراءات فى متن بحثه (جاى وميلز وإيراسيان ، ٢٠١٢ :٢٥١ ، ٢٥٩) .

وعند تطوير مقياس جديد ، يتوقع الباحثون الذين يتبعون إجراءات صارمة لبناء المقياس توفير معلومات وافية وكافية عن صدق المقياس . وعلى الرغم من الأخذ فى الاعتبار أهمية الصدق المرتبط بالمحك وصدق البناء لأداة القياس الجديدة تبقى أيضاً المعلومات المتعلقة بصدق المحتوى تحتاج إلى مراجعة لضرورتها لاشتقاق استنتاجات عن جودة المقياس . كما يظل يُنظر إلى هذه الأنواع على أنها مظاهر صدق مترابطة ، وليست مستقلة (جاى وميلز وإيراسيان ، ٢٠١٢ :٢٥١؛ Polit & Beck, 2006: 489).

ويُعد صدق المحتوى مفيداً للباحثين من مستخدمى أدوات القياس المعدة من قبل آخرين ، ويكشف بفعالية عن مظهر هام من جودة درجات الاختبار (Sireci & Faulkner-Bond, 2014: 101) .

وقد أكد الأدب السيكولوجى على أهمية اتباع عملية صارمة للتحقق من صدق المحتوى لأداة القياس ، باعتباره أمراً بالغ الأهمية للتنبؤ بكفاءة الأداة لأجل تقليل أو حذف أخطاء القياس التى ربما تحدث عندما تكون القياسات المتعددة أمراً متطلباً ، ويسمح للأداة بأن تشمل جميع جوانب البنية⊗ (Haynes,Richard& Kubany,1995:239; Sangoseni,Hellman & Hill, 2013: 2)

وتحليل صدق المحتوى فى أى عملية لبناء أو تعديل الاختبار يُعد أمراً مهماً ، فدليل صدق المحتوى لا يساعد فى تعريف مفاهيمى للبنية موضع الاهتمام ، ولكنه يُعد أيضاً أساساً للتفسير الصحيح لتباين الدرجات التى يتم الحصول عليها ، ومع ذلك ، فهذا الدليل نادراً ما يتم

⊗ البنية Construct هى الخاصية النظرية Theoretical Attribute التى يقيسها الاختبار .

الحصول عليه وتقديمه بشكل تفصيلى حتى على الرغم من أنه متطلب بشكل كبير

(Haynes, Richard & Kubany, 1995; Carretero-Dios & Pérez, 2007)

وقد راجع Yaghmaie (2003: 27) (٣٨) بحث منشور فى دورية International Journal of Nursing Studies خلال العامين (١٩٩٥-١٩٩٦) ووجد أن (٢٠ %) فقط من الباحثين سجلوا عملية صدق المحتوى فى بحوثهم ، بينما (٨٠ %) منهم لا يوثقون هذا النوع من الصدق . وأكد على أن قياس وتسجيل صدق المحتوى لأدوات القياس أمر مهم على أساس أن نتائج هذا النوع من الصدق يمكن أن تُسهم فى دعم وضمان صدق البناء ، وتعطى الثقة للقارىء والباحث فى الأدوات حيث يمكنه فهم عملية قياس صدق المحتوى التى تُمكن بدورها من الوصول إلى تفسيرات دقيقة للنتائج .

وينبع التأكيد على أن صدق المحتوى مكون مهم لتقييم صدق الاستنتاجات المشتقة من درجات الاختبار من أنه يوفر دليلاً على صدق البناء لأداة القياس ، ويوفر أساساً لبناء صور متكافئة فى التقييمات واسعة المدى Large-Scale Assessments (Ding & Hershberger, 2002: 283)

ولأن صدق المحتوى متطلباً قبلياً لصدق البناء والصدق المرتبط بالمحك فمن المهم أن يحظى على الأولوية القصوى أثناء بناء أدوات القياس ، لأهميته للباحثين الذين ينشدون بناء أدوات قياس عالية الجودة ، فالأداة التى تفتقد صدق المحتوى يصعب جداً إثبات ثباتها (Polit,Beck&Owen,2007:459;Bayrak&Yurdugul,2012:264;Zamanzadeh,etal. 2014: 163) .

ويذكر فاروق عبد الفتاح موسى (١٩٩٠ : ١٨١) أن صدق المحتوى يُعتبر من المكونات الهامة فى عملية حساب الصدق للاختبارات النفسية والتربوية . ومن الصعب تصور اختباراً صادقاً يفتقد صدق المحتوى ، وأن أفضل طريقة لمصمم الاختبار ولمن يستخدمه للحكم على صدق المحتوى هى فحص العلاقة بين بنود الاختبار ومجال محتوى الاختبار .

ونظراً لندرة الدراسات والبحوث التى درست تكمية صدق المحتوى من جهة وتناثر الصيغ الإحصائية المستخدمة فى التكمية فى الأدب السيكولوجى هنا وهناك ، أدى ذلك على سبيل المثال إلى أن يذكر Janda (1998: 82) "أنه على خلاف الثبات ، لا يمكن وصف صدق المحتوى بواسطة إحصاء منفرد Single Statistic ، وأن أفضل ما يمكن القيام به بناء صدق المحتوى وبعد ذلك يتم اتباع الإجراءات الإمبريقية لتوفير دليل غير مباشر لصدق محتوى الاختبار ، وأن أفضل سيناريو أن يذكر مُعد الاختبار .. "يبدو صدق محتوى الاختبار جيداً" ، والسيناريو الأسوأ أن يذكر مراجع الاختبار .. "يبدو صدق محتوى الاختبار مشكوك فيه" ، وأنه من غير الممكن الوصول إلى استنتاجات مطلقة بشأن إلى أى مدى يُظهر اختبار ما صدق المحتوى".

وفى نفس السياق يتفق عزت عبد الحميد (٢٠١١ : ٥١٢) وجاى وميلز وإيراسيان (٢٠١٢ :٢٥٣) على أن صدق المحتوى يتحدد من خلال أحكام الخبراء (أى تصديق المحتوى) ، ولا توجد صيغة أو إحصاء يمكن من خلالهما حسابه ، ولا توجد طريقة للتعبير عنه كمياً أى لا يتم التعبير بصورة رقمية عن أدلة صدق المحتوى .

وفى نفس السياق أيضاً ، واعتماداً على توجه صدق المحتوى نحو الاختبارات التحصيلية يذكر (Kubiszyn&Borich, 1990:283-284) أن إجراءات تحديد صدق المحتوى لا تنتج معاملات للصدق ، حيث يتم إثبات صدق المحتوى بمقارنة بنود الاختبار بالأهداف التعليمية لتحديد أياً

من البنود تطابق أو تقيس الأهداف ، وبعد الإنتهاء من هذا الفحص يتم التحكيم على صدق المحتوى ، ولا يتم حساب معاملات ارتباط وبدلاً منه يُعتمد على أحكام الخبراء .

كما يُشير ميخائيل أسعد (١٩٩٠ : ٤٢٥ – ٤٢٦) إلى أن طبيعة إجراء صدق المحتوى تدل على كونه فعالية ذاتية تقوم على مجموعة من الأحكام الذاتية أو الفردية لتصور المحكمين ، ويمكن إضفاء صفة الموضوعية على معيار صدق المحتوى ، باستشارة المختصين والأخذ بتوجيهاتهم ، غير أن الموضوعية التى تبلغها تبقى دون أساس كمى يدل عليها وعلى درجتها .

ومن الوجهة الكلاسيكية ينظر إلى صدق المحتوى كتعريف نظرى ، ومفهوم كيفى ، وكينونة يندر التحقق منها إمبريقياً ، وأنه مستقل ومنفصل عن الأنواع الأخرى وهى صدق البناء والصدق المرتبط بالمحك التى يمكن فحصها والتحقق منها إمبريقياً (Ding & Hershberger, 2002: 284) .

ويذكر Carmines and Zeller (1979: 22) أنه لا يوجد إتفاق على محك لتحديد إلى مدى تحقيق أداة القياس صدق المحتوى ، بالإضافة إلى غياب قياس موضوعى وصارم Rigorous لتحقيق صدق المحتوى.

وينصح Popham (1990: 101) من يُجمعون دليلاً لصدق المحتوى أن يسعوا لتسجيل نتائجهم كمياً كلما أمكن ، مثل توثيق عدد المحكمين ، وعدد الفرص المتاحة للمحكمين لمراجعة محتوى الاختبار ، ونسبة المحكمين الموافقين أو المؤيدين لكل وحدة من وحدات الاختبار، بالإضافة إلى وصف أية ظروف مهمة مرتبطة بتوصيات المحكمين .

وأوصت نتائج دراسة (Jingcheng, Xiankun and Zhenqiu (2012 بأن يتم توثيق خصائص ومؤهلات الخبراء وعملية تقويم صدق المحتوى والنتائج الرئيسة فى متن البحوث .

ولاحظ الباحث الحالى – حسب ما توافر لديه من مصادر – أن أدب القياس النفسى لا يحوى الكثير بشأن نفع القياسات الخاصة بتكمية آراء المحكمين فى سياق صدق المحتوى ، وأن هناك ندرة فى المعلومات المتصلة بدقة هذه الصيغ الإحصائية ، ويُعد ذلك أحد مبررات التصدى للدراسة الحالية .

ويتفق (Ding&Hershberger, 2002: 284; Bayrak &Yurdugul,2012: 264) على أن تقويم صدق المحتوى يُعد مسعى يحمل العديد من التحديات لوجود عدد قليل من الدراسات الامبريقية التى استخدمت طرقاً كمية لتقييم صدق المحتوى اعتمدت على تحويل البيانات الكيفية للمحكمين إلى الشكل الكمى ، ويُعد ذلك مبرراً آخر للتصدى للدراسة الحالية .

ومن جهة أخرى اقتصر الجهد البحثى فى دراسات الصدق فى البيئة المصرية – على حد معرفة الباحث – على دراسة سامية سمير شحاتة (٢٠٠٣) التى أوصت بوضع معادلة إحصائية لتكمية استجابات المحكمين ، كما اقتصر الجهد البحثى فى البيئة العربية على دراسة عبد الله على القاطعى (١٩٩٣) التى استكشفت العلاقة بين صدق استجابات المحكمين والمفهوم الإحصائى لصدق البنود ، ويُعد ذلك مبرراً إضافياً للتصدى للدراسة الحالية .

ومن جهة أخرى – لاحظ الباحث الحالى – أن الممارسات البحثية الراهنة المتعلقة بإجراءات التحكيم على بنود الاختبارات والمقاييس لا تعدو فى أغلب المواقف كونها روتينية وغالباً تنطوى على فحص الصياغة اللغوية للبنود والارتباط بمجال المحتوى دون الانطلاق نحو تعدد المحكات الوارد فى أدب القياس النفسى . إن عدم تحديد المحكات

يحولها إلى عوامل تهدد صدق الاختبار مثل غموض وإرباك البنود الاختبارية وتعقد بنية الجمل .

وتُخضع الدراسة الحالية خمسة قياسات لتكمية صدق المحتوى للفحص الإمبريقى باستخدام محكات الوضوح والبساطة والغموض والارتباط باستخدام ثلاث أدوات (مقياس الهزيمة النفسية ، مقياس الملل الأكاديمى ، استبيان الأسلوب المعرفى اللفظى–البصرى) ، وتتحدد مشكلة الدراسة الحالية فى محاولة الإجابة عن الأسئلة التالية :

هل تختلف تقديرات الصيغ الإحصائية الأربع (مؤشر صدق المحتوى CVI ، ومعامل كابا k ، ونسبة صدق المحكمين CVR ، المؤشر r_{WG}) لتكمية صدق المحتوى للمقاييس الثلاثة على مستوى البنود ؟

هل تختلف تقديرات الصيغ الإحصائية الخمس (مؤشر صدق المحتوى CVI ، ومعامل كابا k ، ونسبة صدق المحكمين CVR، ومتوسط نسبة التطابق ACP ، المؤشر r_{WG}) لتكمية صدق المحتوى للمقاييس الثلاثة على مستوى المقياس ؟

أهداف الدراسة

تهدف الدراسة الحالية إلى :

مقارنة أداء الصيغ الإحصائية الأربع (مؤشر صدق المحتوى CVI ، معامل كابا k ، نسبة صدق المحكمين CVR ، المؤشر r_{WG}) لتكمية صدق المحتوى للمقاييس الثلاثة على مستوى البنود .

مقارنة أداء الصيغ الإحصائية الخمس (مؤشر صدق المحتوى CVI ، ومعامل كابا k ، ونسبة صدق المحكمين CVR ، ومتوسط نسبة التطابق ACP ، المؤشر r_{WG}) لتكمية صدق المحتوى للمقاييس الثلاثة على مستوى المقياس .

أهمية الدراسة

الأهمية النظرية

تتحدد الأهمية النظرية فى التالى :

تتصدى لدراسة موضوع الصدق الذى يُعد أحد أهم الموضوعات فى القياس النفسى والتربوى والاجتماعى .

تُسهم فى إثراء الأدب السيكولوجى والتربوى الخاص بدراسات الصدق وعلى وجه التحديد صدق المحتوى .

تقدم إطار نظرى واف عن خمس صيغ إحصائية لتكمية صدق المحتوى (مؤشر صدق المحتوى CVI ، معامل كابا k ، نسبة صدق المحكمين CVR ، متوسط نسبة التطابق ACP ، المؤشر r_{WG}) .

الأهمية التطبيقية

تتحدد الأهمية التطبيقية فى التالى :

تختبر أداء أربع صيغ إحصائية لتكمية صدق المحتوى امبريقياً على مستوى البنود .

تختبر أداء خمس صيغ إحصائية لتكمية صدق المحتوى امبريقياً على مستوى المقياس .

مصطلحات الدراسة

صدق المحتوى

يُشير إلى درجة مطابقة محتوى الاختبار مع أغراض القياس (Sireci & Faulkner-Bond, 2014: 101) . أو درجة احتواء أداة على عينة مناسبة من البنود للبنية المقاسة (Polit & Beck, 2004: 423) . ويُعرف أيضاً بالمدى الذى تمثل فيه الأداة بشكل كاف المجال البحثى موضع الاهتمام عند محاولة قياس الظاهرة (Waltz, Strickland & Lenz, 2005: 155) .

تكمية صدق المحتوى

يُعرفه الباحث الحالى إجرائياً بأنه استخدام واحداً أو أكثر من المؤشرات أو المعاملات الإحصائية لتحويل الاستجابات الكيفية لمجموعة المحكمين إلى شكل كمى على مستوى البند Item-Level أو على مستوى المقياس ككل Scale-Level .

الاطار النظرى للدراسة

الصدق Validity

يُقصد به قدرة الأداة على قياس خصائص التكوين أو البنية محل الدراسة ، وهو عامل حيوى فى اختيار وتطبيق الأداة ، والصدق ليس خاصية للأداة ولكنه خاصية للدرجات التى يتم الحصول عليها من استخدام الأداة لغرض محدد مع مجموعة معينة من المستجيبين ، وبناء على ذلك يجب الحصول على دليل للصدق فى كل دراسة تستخدم فيها الأداة (Zamanzadeh, et al., 2014: 163) .

ويعتمد صدق الاختبار اعتماداً مباشراً على صدق بنوده ، وذلك لأن أى زيادة فى صدق البنود تؤدى إلى زيادة قصد الاختبار . كما تختلف الاختبارات فى مستويات صدقها تبعاً لاقترابها أو ابتعادها من تقدير تلك الصفة التى تهدف إلى قياسها (فؤاد البهى السيد ، ١٩٧٨ : ٤٥٧) .

ويذكر رجاء محمود أبو علام (٢٠٠٧ : ٤٦٥) أن الصدق أهم خاصية من خواص المقياس ، ويشير مفهوم الصدق إلى الاستدلالات الخاصة التى نخرج بها من درجات المقياس من حيث مناسبتها ومعناها وفائدتها . وتحقيق صدق المقياس معناه تجميع الأدلة التى تؤيد مثل هذه الاستدلالات ، ولذلك يشير الصدق إلى مدى صلاحية استخدام درجات المقياس للقيام بتفسيرات معينة .

ويضيف Slavin (1997: 546) أن الصدق هو قياس لدرجة مناسبة (كفاءة) الاختبار للمستهدف من استخدامه ، ويُشير إلى ما إذا كان الاختبار يوفر نوع المعلومات المرجوة . أما Anastasi and Urbina (1997: 113) فيذكرا أن صدق الاختبار يتعلق بمدى صلاحية الاختبار وصحته فى قياس ما نُوى أنه يقيسه ، ويدل على أمرين هما : ما الذى يقيسه الاختبار؟ وكيف ينجح فى قياسه؟ أى يدلنا على ما يمكن استدلاله من درجات الاختبار ، ويحذرا من قبول اسم الاختبار كمؤشر لما يقيسه الاختبار .

وقد تم تصنيف طرق تحقيق الصدق بواسطة الجمعية الأمريكية لعلم النفس فى دليل النشر الصادر عام ١٩٥٤ حيث تم تقسيم الصدق إلى أربعة أنواع هى : صدق المحتوى Content Validity ، والصدق التنبؤى Predictive Validity ، والصدق التلازمى Concurrent Validity ، وصدق البناء أو التكوين الفرضى Construct Validity ، ثم عدلت الجمعية هذا التصنيف إلى ثلاث فئات فى عام ١٩٦٦ حيث تم دمج الصدق التلازمى والتنبؤى ضمن فئة واحدة هى الصدق المرتبط بالمحك Criterion-RelatedValidity (حيدر إبراهيم ظاظا ، ٢٠١١ : ٢٤٠٠)

ويشتمل دليل "معايير القياس النفسى والتربوى – ١٩٩٩م" الذى طورته الجمعية الأمريكية للبحث التربوى AERA ، والجمعية الأمريكية لعلم النفس APA ، والمجلس القومى للقياس فى التربية NCME والتى رُوجعت حديثاً عام ٢٠١٢م ، قائمة شاملة لمعايير الصدق عددها ٢٤ معيار ، بحيث إذا تحققت ، فإنها تسمح للباحثين بالتوصل إلى مزاعم (ادعاءات) مقبولة فيما يتعلق بالتفسيرات المحددة بالسياق ، وبالنسبة للباحثين المبتدئين تسمح لهم بمناقشة شاملة لجميع هذه المعايير . وتُعرف الصدق بأنه يرجع إلى "الدرجة التى تدعم فيها الأدلة والنظرية

التفسيرات لنتائج الاختبارات التى تنطوى عليها التطبيقات المقترحة للاختبارات (AERA, APA & NCME, 1999: 9).

ويُعد التصور الذى اقترحه Messick (1989: 18) أحد التصورات الحديثة فى إطار نظرية الصدق الحديثة ، حيث استقر تصور الصدق على التركيز على معنى درجات الاختبار أو قدرته التفسيرية . بمعنى أن الاختبار يكون صادقاً عندما تكون الاستدلالات المشتقة منه مناسبة ولها معنى ومفيدة ، ومن ثم فدرجة الاختبار يكون لها مغزى وتكون مفيدة طالما تمكن مستخدم الاختبار من أن يشتق منها استدلالات بناءً على تعليمات الاختبار .

ووفقاً لنظرية الصدق الحديثة ، فإنه ليس هناك اختبار صادق ، ولكن يطلق الصدق على الاستنتاجات أو تفسير نتائج الاختبار . فالصدق خاصية للاستدلال ، وليس للأداة ، ويكون الاختلاف فى صدق التفسيرات من حيث الدرجة . ويُفضل أن ينظر للصدق على أنه فرضية أو جدل تفسيرى تُجمع من أجله الأدلة والبراهين لدعم الاستنتاجات المقترحة (إقبال زين العابدين درندرى ، ٢٠١٢ : ٤٣) .

صدق المحتوى Content Validity

يُعرف أيضاً بالصدق المرتبط بالمحتوى Content-Related Validity ، والصدق الداخلى Intrinsic Validity ، وصدق الارتباط Relevance Validity ، والصدق المنطقى Logical Validity ، وصدق التمثيل Representative Validity ، وصدق المعاينة Sampling Validity ، وصدق التعريف Definition Validity (Yaghmaie, 2003: 25; Zamanzadeh et al., 2014: 163-164) .

ويحدد صدق المحتوى تعريف الصفة النفسية المقاسة ، ويُعد أول معايير تحديد صدق الأداة وأبسطها وأكثرها سهولة وتوليداً للثقة أو اليقين

المباشر من أن الأداة تقيس ما نُوى لها ومنها أن تقيس (ميخائيل أسعد ، ١٩٩٠ : ٤٢٢) .

ويعنى الفحص المنظم لمحتوى الاختبار (أى بنوده) ، لتحديد ما إذا كان يشتمل على عينة ممثلة لمجال السلوك موضع المقاس أم لا (Anastasi & Urbina, 1997: 113) . ويدل على مدى تمثيل بنود الاختبار لمحتوى الخاصية (السمة / القدرة) المراد قياسها والسابق تحديدها (أمين على محمد سليمان ، ٢٠١٠ : ٥٨٥) ، كما يدل على درجة تمثيل بنود الاختبار للسمة موضع القياس (حيدر إبراهيم ظاظا ، ٢٠١١ : ٢٤٠٠) .

ويُشير صدق المحتوى أيضاً إلى قدرة مجموعة منتقاة من البنود على أن تعكس خصائص البنية المقاسة ، ويوفر دليلاً تمهيدياً على صدق البناء للأداة ، بالإضافة إلى أنه يوفر معلومات عن تمثيل البنود ووضوحها ، ويساعد على تحسين الأداة من خلال تحقيق توصيات مجموعة المحكمين (Zamanzadeh et al., 2014: 164) .

دور التحكيم فى صدق المحتوى

يُعتبر استخدام تقديرات المحكمين جزء أساسى من بناء الاختبارات والمقاييس النفسية بحيث لا يُعد الاختبار كفء إلا إذا أضاف مجموعة من المتخصصين آرائهم إلى تجارب مصمم الاختبار ليؤكدوا صدق الاختبار .

وفى ضوئه يتم اختيار مجموعة من الخبراء المشهود بكفاءتهم فى المجال ويعرض عليهم الاختبار فى مسودته وأمام كل بند من بنوده يضع كل خبير علامة تدل على أن البند يقيس أو لا يقيس ما وضع لقياسه كما تترك له الحرية إذا كان غير متأكد من أن البند يقيس أو لا يقيس ما وضع لقياسه بأن يضع علامة تدل على تأكده

(محمد إسماعيل عمران وعبد الله السيد عبد الجواد وفهد عبد الله الدليم ، ١٩٩٤ : ٧٦-٧٧ ؛ رمزية الغريب ، ١٩٩٦ : ٦٩٠) .

ومن المشكلات الشائعة التى تواجه الباحث السيكولوجى فى استخدامه لتقديرات المحكمين : كيف يمكن الثقة فى آراء المحكمين؟ وما الضوابط الإجرائية التى يمكن أن تزيد بدورها من فاعلية تقديرات المحكمين ؟ (سامية سمير شحاتة ، ٢٠٠٣) .

كما أن الباحث حال استطلاعه لرأى المحكمين فإنه يعتبره تحصيل حاصل حيث أنه يعتمد على أنواع أخرى من أدلة (مؤشرات) الصدق ليتحقق من صدق اختباره . كذلك قد نجد تبايناً كبيراً فى تقديرات المحكمين لصدق الاختبار ، كما أن الباحث مصمم الاختبار عندما يقدم اختباره لمجموعة من المحكمين فإنه يقدم لهم تعريفاً إجرائياً للسمة التى يقيسها الاختبار والبنود التى توجد فى الاختبار ليرى هل ينطبق هذا التعريف مع البنود حتى يمكن الحكم على أن هذا الاختبار صادق أم لا؟ من هنا نرى هؤلاء المحكمين يفحصون البنود للتعرف على مدى انطباق البنود مع التعريف المطروح أم لا؟ أى أن المحكمين سيصدقون على التعريف الإجرائى الذى وضعه مصمم الاختبار للسمة المراد قياسها بصرف النظر عما إذا كان هذا التعريف ينطبق مع الخصائص المميزة للسمة أم لا ؟ .

ومن الممارسات الروتينية أن يعرض مصمم الاختبار اختباره على محكمين معروفين له مسبقاً أو معروفين للمشرف على البحث دون النظر إلى ما إذا كان هؤلاء المحكمين متخصصين فى هذا الفرع من المجال أم لا؟ ، كما أن الانتهاء من عملية التحكيم قد تكون عملية شاقة إذا أخذ فى الاعتبار ردود فعل بعض المحكمين التى تتسم غالباً باللامبالاة وعدم الاهتمام ، الأمر الذى يدفعهم فى بعض الأحيان إلى مجرد وضع

علامات بالموافقة دون بذل أى مجهود منهم لفحص مضمون العبارات ومدى مطابقتها لمضمون الخاصية أو السمة موضع القياس .

أما بالنسبة لعدد المحكمين فيذكر Lynn (1986: 383) أن وجود خمس محكمين يمكن أن يوفر مستوى كاف لضبط اتفاق الصدفة Chance Agreement ، وفى حالة الاقتضاء يمكن تخفيض العدد إلى ثلاثة محكمين ، أما تعيين أقل من ثلاثة محكمين فينطوى على مخاطرة كبيرة فى الاستنتاج الخاطىء بشأن استيفاء أو عدم استيفاء صدق محتوى أداة القياس .

ويذكر McCain (1984) أن البنود التى لا تحقق الحد الأدنى من الاتفاق المتطلب يجب حذفها أو تعديلها ، وهذا يستدعى إعادة تقويم أداة القياس لأجل الحصول على صدق محتوى كاف ، وفى هذه الحالة على مطور أداة القياس تحديد إذا ما كان سيستعين بنفس المحكمين أو محكمين جدد لهم نفس الكفاءة ، وإعادة التحكيم على البنود المعدلة يتطلب فاصلاً زمنياً ما بين ١٠ إلى ١٤ يوماً .

محكات التحكيم على بنود أدوات القياس

قدم Davis (1992) خمسة محكات لتقييم كل بند تبعاً للتمثيل Representativeness والارتباط والفهم Comprehension والغموض والوضوح باستخدام مقياس استجابة من نوع ليكرت رباعى التدرج .

كما قدم Waltz and Bausell (1983) وأيضاً Lynn (1986) أربعة محكات لتكمية صدق المحتوى لأدوات القياس سيستخدمها الباحث الحالى وهى :

أولاً : محك الارتباط (الصلة) Relevance

يُكلف المحكمون لتقدير ارتباط أو صلة كل بند بالمتغير المنوى قياسه ، وفى هذا المحك تعد معتقدات مُعد الاختبار بشأن قياسات كل بند بمثابة الفرض واستجابات الخبراء هى بيانات للتأكيد أو النفى . وحتى

فى حالة قياس جميع البنود لخاصية أو تكوين منفرد تكون مراجعة الخبير مفيدة ، وتقوم آلية تقويم ارتباط البند على تقديم مُعد الاختبار للتعريف الإجرائى للمتغير المستهدف بالقياس إلى مجموعة المحكمين أو الخبراء (DeVellis, 2003: 86) .

ثانياً : محك الوضوح Clarity

يمكن أن يُقوم المراجعون أيضاً وضوح البند . فربما يكون محتوى البند ذو ارتباط بالمتغير المستهدف لكن صياغته ربما تسبب مشكلة ، وهذا يُحمل على ثبات البند لأن وجود بند غير واضح بدرجة أكبر من بند واضح يمكن أن يعكس عوامل دخيلة Factors Extraneous على المتغير الكامن . وعلى معُد الاختبار أن يطلب من المراجعين فى تعليمات الاختبار أن يشيروا إلى البنود المربكة Awkward أو المحيرة Confusing التى تخلط بين أمرين مختلفين واقتراح صياغات بديلة إذا ما رغبوا فى ذلك (DeVellis, 2003: 86) .

ثالثاً : محك البساطة Simplicity

يمكن أن يُقوم المراجعون بساطة البند ودقته Precision ، أى استخدام جمل دقيقة للتعبير عن بنية البند دون زيادة أو نقصان ، حيث أن تطويل الجملة أو تقصيرها يُخل ببساطتها . حيث تُوضع جملة البند فى حالة التطويل أو الزيادة تحت مُسمى الإطناب/الإسهاب Redundancy/Circumlocutions ، وفى حالة تقصير الجملة أو الإيجاز تُوضع تحت مُسمى الإيجاز Conciseness ، وفى الحالتين فإن جملة البند تفتقر للبساطة والدقة (Delgado-Rico, Carretero-Dios & Ruch, 2012: 453) .

رابعاً : محك الغموض Ambiguity

يمكـــن أن يُقـــوم المراجعـــون غمـــوض صـــياغة البنـــد Ambiguous Item Phrasings من حيث المعنى ، أى أن محتوى البند لا لبس فيه Unequivocally ، ويتصل ذلك بالإبهام Vagueness بما لا يسمح بتفسيرات محددة ومميزة ذات معنى خلال السياق مثل البنود التى لها معنيين محتملين Two Potential Meanings ، ومن مصادر غموض البند احتواء صياغته على كلمات مثل : بعض Some ، كبير Large ، صغير Small ، زمن بعيد Long Time ، عادة Usually ، بشكل منتظم Regularly . وأحياناً ربما يكون الغموض معجمى Lexical Ambiguity أو نحوى (تركيبى) Syntactic أو سيمانتى Semantic ، وبعامة تسهم البنود الغامضة فى سوء التفسيرات Misinterpretations (Delgado-Rico, Carretero-Dios & Ruch, 2012: 453) والبنود الغامضة هى بنود سيئة الإعداد (الصياغة) Poorly Drafted Items لا تستوف غرض التقويم لأنها تنتج استجابات متحيزة وهذا يعنى عدم تغطية مظاهر مرتبطة بالمحتوى تعتبر ذات ارتباط بالتكوين المستهدف بالقياس . والغموض يعنى التحكيم على احتمالات بأن البند يمكن أن يُفسر بطرق مختلفة (Delgado-Rico, Carretero-Dios & Ruch, 2012: 450-453) .

ويوجد تباين واضح جداً فى الأدبيات بشأن جدوى استخدام أكثر من محك وجدوى استخدام القرار ثنائى التقسيم عن طريق اختزال متصل القياس الرتبى رباعى النقاط أو التدرج إلى متصل ثنائى أو وضع تقديرات الخبراء على مقياس تدرج رتبى (DeVellis, 2003: 86) .

ونظـــراً لظهـــور بعـــض التـــداخل فـــى المعنـــى Overlap in the Meaning فى المحكات الأربعة السابقة ، وبخاصة الوضوح والغموض والبساطة ، تمت التوصية باستخدام المقياسين الأول

والثـــــانى وهمـــــا الارتبـــــاط والوضـــــوح فقـــــط (Shrock & Coscarelli, 2007: 200).

مواصفات المحكمين

بمراجعة بعض الأدبيات يمكن إجمال بعض المواصفات المتطلبة فى المحكم على البنود المتضمنة فى أدوات القياس كالتالى :

لديـه القابليـة والاسـتعداد والدافعيـة للمسـاهمة فـى فحـص البنـود (فاروق السيد عثمان ، عبد الهادى السيد عبده ، ١٩٩٥ : ٢٤١) .

يملـــك الخبـــرة الكافيـــة بالبنيـــة المســـتهدفة بالقيـــاس (Grant & Davis, 1997; Sireci, 1998: 1) .

معـروف ويملـك سـمعة محترمـة فـى دراسـة المتغيـر المقـاس (Wynd et al., 2003: 514) .

يتحلـــــى بالأمانـــــة العلميـــــة والموضـــــوعية (صالح سليمان البقعاوى ، ٢٠١٤ : ٢٠٢) .

طـــــرق تكميـــــة صـــــدق المحتـــــوى Quantification Methods of Content Validity

(١) مؤشر صدق المحتوى (CVI) Content Validity Index

يُعد مقياساً أكثر شيوعاً لصدق المحتوى واستخدم لسنوات عديدة ، وقد قدمه (1977) Martuza ، وشرحه (1986: 382-385) Lynn بأنه منهجية أو عملية من خطوتين لتقدير صدق المحتوى . تسمى الخطوة الأولى خطوة البناء وتقوم على تحديد مجال المحتوى من خلال مراجعة شاملة للأدب السيكولوجى متبوعاً ذلك بتوليد بنود أداة القياس . ويكتمل بنـاء الأداة ليتضـمن تعليمـات للمسـتجيبين وآليـة تقـدير الـدرجات أو التصـحيح . أمـا الخطـوة الثانيـة فتسـمى خطـوة التكمـيم/الحُكم Judgment / Quantification وتتضمن تطوير طريقة موضوعية لتكمية صدق المحتوى ، وذلك عندما يقوم فريق مختار من الخبراء بتقويم الأداة

وأثناء هذا التقويم يستخدم الخبراء مقياس تقدير من نوع ليكرت ، ونسبة الخبراء الموافقين على ارتباط البند مثلاً يوفرون قياساً كمياً لصدق المحتوى يسمى مؤشر صدق المحتوى .

ويوجد نوعين من مؤشر صدق المحتوى ، الأول مؤشر صدق المحتوى على مستوى البند (I-CVI) Item-Level ، والثانى مؤشر صدق المحتوى على مستوى المقياس (S-CVI) Scale-Level ويمكن استخدام طريقة حساب المتوسط (S–*CVI* / Ave) (Ave) Averaging Calculation Method أو طريقة حساب الاتفاق الكلى (UA) Universal Agreement Calculation Method (S–*CVI* /*UA*) (Polit & Beck, 2006: 491)

ويستخدم الباحثون معلومات مؤشر صدق المحتوى على مستوى البند I-CVI ليرشدهم فى مراجعة وحذف أو استبدال البنود ، ومع ذلك فإن الباحثين لا يوفرون عادة معلومات عن قيم مؤشر صدق المحتوى على مستوى البند فى تقاريرهم البحثية ، لكنها تُسجل فقط فى دراسات مناهج البحث التى تركز على وصف عملية صدق المحتوى . والذى يسجل غالباً فى دراسات بناء المقاييس هو مؤشر صدق المحتوى على مستوى المقياس S–CVI وهنا تكمن المشكلة (Polit & Beck, 2006: 491)

وقد طرح Waltz and Bausell (1981: 71) قيدين Two Limitations لاستخدام مؤشر صدق المحتوى *CVI* لتقييم صدق المحتوى هما احتمالية تضخم اتفاق الصدفة Possibility of Chance Inflation of Agreement ، واعتماد المؤشر على عدد الفئات المستخدمة فى التقدير .

لذا كانت من أوجه النقد الموجهة لهذا المعامل أنه يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى ، وينصح باستخدام معامل

كابا كبديلاً لأنه يخفض أو يقلل اتفاق الصدفة بين الخبراء Discount for Chance Agreement (Beckstead, 2009: 1274) .

(٢) إحصاء كابا (k) Kappa Statistic

قدمه (1960)Cohen لفحص الاتفاق بين المحكمين Interrater Agreement . ويمثل إحصاء كابا نسبة الاتفاق المتبقية بعد حذف إتفاق الصدفة فضلاً عن مقارنة النسبة الكلية للإتفاق $(P_{\circ})$ Total Proportion of Agreements للقيمة العظمى لـ ١٠٠% ، وتقارن النسبة الكلية بأعلى قيمة ممكنة للإتفاقات بواسطة الصدفة وحدها $(1-P_c)$ ، ليعطى التوزيع الحدى لتقديرات البند التى يتم تعيينها بواسطة كل خبير من فريق المحكمين . حيث (P_c) نسبة الاتفاق المتوقع حدوثها بالصدفة وحدها ، $(P_{\circ}-P_c)$ تمثل الملاحظات الحقيقية مقابل إتفاقات الصدفة (الاتفاقات المتوقع حدوثها بالصدفة) . ويُعطى إحصاء كابا من المعادلة التالية :

$$k=\frac{P_{\circ}-p_c}{1-p_c}$$

ويُحسب احتمال إتفاق الصدفة من المعادلة التالية :

$$P_c=\frac{N!}{A!(N-A)!}*.5^N$$

وهذه المعادلة تعتمد على حساب مضروب العدد الكلى للمحكمين N ، ومضروب عدد المحكمين المتفقين A (Polit, Beck, Owen, 2007: 463) .

وتمتد قيم إحصاء كابا من (−١) إلى (+١) ، وتبين القيم الموجبة لمعامل كابا إتفاق بين المحكمين يحدث بشكل متكرر عما هو متوقع بواسطة الصدفة ، والقيمة (+١) تظهر إتفاق تام أو كامل عبر المقدرين

، والقيمة الصفرية تبين أن الاتفاقات لم يعد متوقع حدوثها على أساس الصدفة ، وتظهر القيم السالبة أن المقدرين يتفقون بشكل أقل تكراراً عما هو متوقع حدوثه بالصدفة ، والقيمة (−١) تبين عدم إتفاق بشكل تام (Suen & Ary, 2014: 114) .

وقد أوصى (1975) Gelfand and Hartmann بأن الحد الأدنى المقبول للإتفاق بين المقدرين يبلغ (٠,٦٠) . وعرضت بعض الأدبيات مقارنة لقوة الاتفاق وقيم معامل كابا كما هو موضح بالجدول التالى :

جدول (١) مقارنة البارامترات الهامشية لمعاملات كابا

دراسة Landis & Koch (1977)		دراسة Cicchetti (1984); Fleiss (1971)	
قوة الاتفاق	إحصاء كابا	قوة الاتفاق	إحصاء كابا
ضعيف Poor	أقل من ٠,٠٠	ضعيف Poor	أقل من ٠,٤٠
قليل Slight	٠,٠٠ – ٠,٢٠	معقول Fair	٠,٤٠ – ٠,٥٩
معقول Fair	٠,٢١ – ٠,٤٠	جيد Good	٠,٦٠ – ٠,٧٤
متوسط Moderate	٠,٤١ – ٠,٦٠	ممتاز Excellent	٠,٧٥ – ١,٠٠
جوهرى Substantial	٠,٦١ – ٠,٨٠		
تام Almost Perfect	٠,٨١ – ١,٠٠		

ويذكر (2003: 510-511) Wynd, et al. أن إحصاء كابا مُكمل مهم Important Supplement لمؤشر صدق المحتوى *CVI* لأنه ينتج مؤشراً لدرجة الاتفاق بعد إتفاق الصدفة على غير مؤشر صدق المحتوى غير المصحح لإتفاق الصدفة ، وأن مؤشر صدق المحتوى يبدد المعلومات باختزال تقديرات الخبراء الرتبية متعددة النقاط إلى فئتين مثل أن يكون محتوى البند (مرتبط / غير مرتبط) بمجال القياس.

(٣) نسبة صدق المحتوى Content Validity Ratio (CVR)

قدم Lawshe (1975) معادلة نسبة صدق المحتوى فى الصورة التالية :

$$CVR = \frac{(n_e - \frac{N}{2})}{\frac{N}{2}}$$

حيث N العدد الكلى للمقدرين ، n_e عدد المقدرين الذين يحُكموا على أن البند ذو ارتباط بالمتغير المستهدف بالقياس . ويمكن إعادة التعبير عن نسبة صدق المحتوى CVR باستخدام نسبة المقدرين المؤيدين P_e التى يُعبر عنها بالمعادلة التالية :

$$P_e = \frac{n_e}{N}$$
$$or$$
$$n_e = NP_e$$

وهذه القيمة يمكن التعويض عنها لتصبح نسبة صدق المحتوى CVR فى الصورة التالية :

$$CVR = \frac{(n_e - \frac{N}{2})}{\frac{N}{2}} = \frac{(NP_e - \frac{N}{2})}{\frac{N}{2}} = \frac{\frac{N}{2}(2P_e - 1)}{\frac{N}{2}} = 2P_e - 1$$

والدلالة الإحصائية لقيمة نسبة صدق المحتوى CVR المتحصل عليها تتحدد بواسطة مقارنة القيم الحرجة التى قدمها Schipper والمشتقة اعتماداً على توزيع ذات الحدين Binomial Distribution لصالح دراسة Lawshe (1975) كما بالجدول التالى :

جدول (٢) القيم الصغرى لنسبة صدق المحتوى

عدد المحكمين	القيمة الصغرى	عدد المحكمين	القيمة الصغرى	عدد المحكمين	القيمة الصغرى	عدد المحكمين	القيمة الصغرى
٥	٠,٩٩	٩	٠,٧٨	١٣	٠,٥٤	٢٥	٠,٣٧
٦	٠,٩٩	١٠	٠,٦٢	١٤	٠,٥١	٣٠	٠,٣٣
٧	٠,٩٩	١١	٠,٥٩	١٥	٠,٤٩	٣٥	٠,٣١
٨	٠,٧٥	١٢	٠,٥٦	٢٠	٠,٤٢	٤٠	٠,٢٩

ولأن قيمة الاحتمال P_e تأخذ المدى من (صفر إلى ١) فإن قيم نسبة صدق المحتوى CVR تقع فى الفترة $-1.0 \leq CVR \leq 1.0$ ، وفى حالة مقياس متعدد البند Multi-Item Scale يتكون من J من البنود يتم حساب مؤشر صدق المحتوى (CVI) على مستوى المقياس بأخذ متوسط قيم نسبة صدق المحتوى CVR للبنود التى يتضمنها المقياس ، وهذا يؤدى إلى :

$$CVI = 2\bar{P}_e - 1$$

حيث $\bar{P}_e$ متوسط نسبة تأييد البند Average Proportion of Item Endorsement .

وهذه الطريقة تستخدم على نطاق واسع لتثبت وتكمم صدق المحتوى فى مجالات متنوعة تشمل الرعاية الصحية ، والتربية ، والنمو التنظيمى ، وعلم نفس الشخصية ، وبحوث التسويق (Wilson, Pan & Schumsky, 2012: 199-200) .

وقد قدم Wilson, Pan, and Schumsky (2012) إعادة لحساب القيم الحرجة لنسبة صدق المحتوى CVR ، كما قدم Ayre and Scally (2014) مراجعة لطريقة الحساب الأصلية لنسبة صدق المحتوى من خلال القيم الحرجة .

(٤) متوسط نسبة التطابق Average of Congruency Percentage (ACP)

قدمه Popham (1978) ، ويعتمد على تعيين فريق من الخبراء لتحديد أياً من البنود المتضمنة فى مقياس متطابقة Congruent أو مرتبطة مع بنيته . ثم يلى ذلك حساب نسبة البنود التى يرى الخبير أنها مرتبطة ، ثم يؤخذ المتوسط الحسابى للنسب عبر كل الخبراء .

وقد حدد Waltz, Strickland, and Lenz (2005: 178) نسبة (٩٠%) أو أكثر كمحك معيارى للقبول على مستوى المقياس لأنه أكثر تحرراً Liberal فى تعريفه للتطابق . ويُؤخذ على هذا الإحصائى أنه يكمم صدق المحتوى على مستوى المقياس ككل وليس على مستوى البند.

(٥) المؤشرات $r_{WG}, r_{WG(J)}$

قدم James, Demaree and Wolf (1984: 85-98) صيغة لحساب هذا المؤشر كالتالى :

$$r_{WG} = 1 - \frac{S_X^2}{S_{EU}^2}$$

حيث S_X^2 هو تباين تقديرات المحكمين ، S_{EU}^2 هو تباين الخطأ الموزع المنتظم (الموحد أو المتماثل) Uniform ويُحسب من المعادلة التالية :

$$S_{EU}^2 = \frac{(A^2 - 1)}{12}$$

حيث A عدد بدائل الاستجابة .

ومثل نسبة صدق المحتوى CVR فى حالة البنود ثنائية التقسيم Dichotomous Items أى $A = 2$ يمكن التعبير عن المؤشر r_{WG} بصيغة احتمال التأييد Probability of Endorsement عندئذ يكون :

$$S_X^2 = P_e(1-P_e)$$

$$S_{EU}^2 = \frac{(2^2-1)}{12} = \frac{3}{12} = .25$$

ويكون المؤشر r_{WG} فى الصيغة التالية :

$$r_{WG} = 1 - \frac{P_e(1-P_e)}{.25} = 1 - \frac{4P_e(1-P_e)}{1} = 4P_e^2 - 4P_e + 1 = (2P_e - 1)^2$$

ولأن P_e تأخذ المدى من (صفر إلى ١) فإن قيم المؤشر r_{WG} تقع فى الفترة $0.0 \le r_{WG} \le 1.0$ ، أما فى حالة البنود متعددة الاستجابة Polychotomous (الحالة $A \ge 3$) فإن الحد الأعلى يظل $r_{WG} = 1.0$ ، ولكن القيم السالبة يتم الحصول عليها عندما يكون تباين التقديرات أكبر من المرتبطة بالتوزيع المنتظم وفى هذه الحالات أُوصى أن تُعين قيمة المؤشر r_{WG} بالصفر وهذا يحافظ على الفترة $0.0 \le r_{WG} \le 1.0$ فى حالة البنود متعددة الاستجابة أيضاً .

وقد تم تمديد منطق تقييم الاتفاق لتقديرات مستهدف منفرد Single Target على بند منفرد إلى تقديرات مستهدف منفرد على مقياس متعدد البنود Multi-Item Scale . والمؤشر متعدد البند يمكن حسابه من الصيغة التالية :

$$r_{WG(J)} = \frac{J\left(1 - \frac{\bar{S}_X^2}{S_{EU}^2}\right)}{J\left(1 - \frac{\bar{S}_X^2}{S_{EU}^2}\right) + \frac{\bar{S}_X^2}{S_{EU}^2}}$$

حيث J عدد البنود فى المقياس ، $\bar{S}_X^2$ هو متوسط تباين البنود المتضمنة فى المقياس ، S_{EU}^2 هو تباين التوزيع المنتظم ، وفى حالة $(J=1)$ تؤول صيغة المؤشر $r_{WG(J)}$ إلى المؤشر r_{WG} أى يكون $r_{WG(1)} = r_{WG}$ (Lindell & Brandt, 1997: 273-274) .

دراسات وبحوث سابقة

أجرى Lindell and Brandt (1999) دراسة هدفت إلى مقارنة أربع صيغ إحصائية لتقدير الاتفاق بين المقدرين هى (نسبة صدق المحتوى CVR ، T ، $r^*_{WG(J)}$ ، $r_{WG(J)}$) وبلغ حجم عينة الخبراء (١٥) خبيراً ، ومن بين ما توصلت إليه نتائج الدراسة أن الصيغ الإحصائية الأربع أظهرت اتساقاً فى الأداء فيما بينها وأن المعامل $r^*_{WG(J)}$ هو الأبسط فى الحساب والأكثر قابلية للتطبيق مقارنة بالصيغ الأخرى .

وهدفت دراسة Wynd, et al., (2003) إلى مقارنة طريقتى مؤشر صدق المحتوى ومعامل كابا لتقدير صدق المحتوى لـ Osteoporosis Risk Assessment Tool (ORAT) وقد تم جمع البيانات من ثمان محكمين ، ومن بين ما توصلت إليه نتائج الدراسة تباين نتائج الطريقتين فى فرز البنود المقبولة حيث بين مؤشر صدق المحتوى CVI وجود بند واحد لم يحقق نسبة إتفاق بين المحكمين (CVI = 0.57) ، أما قيم معامل كابا فقد كان منخفضة لسبعة بنود من إجمالى عدد البنود وتراوحت بين (٠,٢٩) إلى (٠,٤٨) ليظهر اتفاقاً منخفضاً بين المحكمين ، وأوصت الدراسة بحذف أو مراجعة البنود الثمانية .

ومن بين ما هدفت إليه دراسة Fung (2010) استخدام ومقارنة ثلاث طرق (مؤشر صدق المحتوى CVI ومعامل صدق المحتوى (VIk) ونسبة صدق المحتوى (CVR) للتأكد من صدق المحتوى لـ Asperger's Syndrome Self-Screening Instrument ، وبلغ حجم عينة المحكمين تسعة خبراء ، ومن بين ما توصلت إليه نتائج الدراسة صلاحية الطرق الثلاث لانتقاء البنود المقبولة ، وقد أوصت الدراسة

باستخدام صدق المحتوى تحت ظروف محددة ومنها صلاحية فريق المحكمين .

فى حين هدفت دراسة Jingcheng, Xiankun and Zhenqiu (2012) إلى مقارنة أداء مؤشر صدق المحتوى على مستوى البند ($I-CVI$) والمقياس ($S-CVI$) وإحصاء كابا المعدل k^* ، ومن بين ما توصلت إليه نتائج الدراسة أن المقياس الذى يحظى بصدق محتوى متميز يكون تقدير مؤشر صدق المحتوى على مستوى البند ($I-CVI \geq .78$) وتقدير المؤشر باستخدام طريقة حساب الاتفاق الكلى ($S-CVI/UA \geq .8$) وتقدير المؤشر بطريقة حساب المتوسط ($S-CVI/Ave \geq .9$) .

أما دراسة Wilson, Pan, and Schumsky (2012) فقد هدفت إلى تمديد جداول القيم الحرجة الأصلية لنسبة صدق المحتوى CVR لـ Lawshe (1975) التى قدمها Schipper ، ومن بين ما توصلت إليه نتائج الدراسة وجود أكثر من قيمة شاذة Anomalous Values فى جدول القيم الحرجة الأصلى ، وأن Schipper استخدم كل من الحساب المتقطع Discrete Calculation والتقريبات المعيارية Normal Approximations لتوزيع ذات الحدين Binomial Distribution لإعداد جداول تجريبية للقيم الحرجة ، كما توصلت الدراسة أن جداول القيم الحرجة الأصلية لنسبة صدق المحتوى تمثل محكاً أكثر تحفظاً لاستبقاء البند .

كما هدفت دراسة Ayre and Scally (2014) إلى مراجعة جداول القيم الحرجة الأصلية لنسبة صدق المحتوى CVR لـ Lawshe (1975) التى قدمها Schipper وجداول القيم الحرجة المعدلة فى دراسة (Wilson, Pan & Schumsky, 2012) باستخدام احتمالات ذات الحدين المضبوطة Exact Binomial Probabilities ، وبلغ حجم عينة المحكمين (٥-٤٠) محكماً ، ومن بين ما توصلت إليه نتائج الدراسة تساوى القيم الحرجة

الأصلية مع المنتجة فى حالة إذا كان حجم عينة المحكمين أقل من (١٠) .

تعقيب على الدراسات والبحوث السابقة

ندرة الدراسات التى تناولت تكمية صدق المحكمين ، حيث لم يسفر البحث فى الأدب السيكولوجى عن أى دراسة فى البيئة المصرية أو العربية .

عدد المحكمين يمكن أن يكون فردياً أو زوجياً .

اتجهت دراستى (Wilson,Pan&Schumsky,2012;Ayre&Scally, 2014) وهما من أحدث الدراسات إلى تطوير نسبة صدق المحتوى CVR .

مؤشر صدق المحتوى CVI هو المقياس الإحصائى الأكثر استخداماً فى الدراسات والبحوث السابقة .

المؤشرات (T ، $r^{*}_{WG(J)}$ ، $r_{WG(J)}$) هى الأقل استخداماً لتكمية محكات صدق المحتوى نظراً للتعقيدات المتعلقة بطرق الاحتساب .

تختلف الدراسة الحالية عن الدراسات والبحوث السابقة كماً وكيفاً فيما يتصل بعدد الصيغ الإحصائية المستخدمة لتكمية صدق المحتوى .

فروض الدراسة

تسعى الدراسة الحالية إلى اختبار الفروض التالية :

لا تختلف تقديرات الصيغ الإحصائية الأربع (مؤشر صدق المحتوى CVI ، معامل كابا k ، نسبة صدق المحكمين CVR ، المؤشر r_{WG}) لتكمية صدق المحتوى للمقاييس الثلاثة على مستوى البنود .

لا تختلف تقديرات الصيغ الإحصائية الخمس (مؤشر صدق المحتوى CVI ، معامل كابا k ، نسبة صدق المحكمين CVR ، متوسط نسبة التطابق ACP ، المؤشر r_{WG}) لتكمية صدق المحتوى للمقاييس الثلاثة على مستوى المقياس .

الطريقة والإجراءات

يتضمن هذا الجزء عرض لعينة الدراسة والأدوات المعدة والمحكات المستخدمة لتقويم صدق المحتوى كالتالى :

أولاً : العينة

بلغ حجم عينة المحكمين• (١٠) محكمين متخصصين فى التربية وعلم النفس بكليتى التربية والآداب بقنا وبيانهم كما بالجدول التالى :

جدول (٣) توزيع عينة المحكمين

الأدوات	التربية				التربية		الآداب		عدد المحكمين
	علم النفس التربوى		الصحة النفسية		المناهج وطرق التدريس		علم النفس		
	أستاذ	أستاذ م.	أستاذ م.	مدرس	أستاذ	مدرس	أستاذ	أستاذ م.	
مقياس الهزيمة النفسية	١	١	٤	٣	–	–	١	–	١٠
مقياس الملل الأكاديمى	١	١	٢	٢	١	١	١	١	١٠
استبيان الأسلوب المعرفى البصرى – اللفظى	١	٢	٣	٢	–	–	١	١	١٠

ثانياً : أدوات الدراسة

(١) مقياس الهزيمة النفسية (إعداد : الباحث)

يعرف الباحث الحالى الهزيمة النفسية Mental Defeat بأنها تصور إحساس الفشل فى الكفاح وفقد المرتبة ، وتصف التأثير الأعمق من الألم على شعور الفرد بذاته ، كما تصف رد الفعل النفسى الشديد للألم الذى يأخذ الفرد بعيداً عن الحكم الذاتى والهوية ، أى تصف إدراك الفرد بأنه فقد الحكم الذاتى كإنسان ، وفقد الرغبة فى ممارسة الضبط أو السيطرة

• ملحوظة : نصح Lynn (1986) أن الحد الأدنى لعدد المحكمين لا يقل عن ثلاثة ، وأن زيادة عدد المحكمين عن عشرة يُعد أمراً غير ضرورى .

والحفاظ على الهوية ، وفقد الاعتقاد بأنه فرد (شخص) لديه الإرادة الحرة.

وقد تم إعداد المقياس بعد الإطلاع على مجموعة من الأدبيات المنشورة ومنها : (Gilbert, 2001a,b ؛ Gilbert & Allan, 1998 ؛ Gilbert, 2007 ؛ Gilbert, Allan, Brough, Melley & Miles, 2002 ؛ Garcia-Campayo, Rodero, Loez del Hoyo, Luciano, Alda & Gili, 2010 ؛ محمد السعيد عبد الجواد أبو حلاوة ، ٢٠١٢ ؛ محمد السعيد أبو حلاوة وراشد مرزوق راشد ، ٢٠١٣ ؛ فضل إبراهيم عبد الصمد ، ٢٠١٣) ، وتتكون الصورة المبدئية من المقياس من (٢٧) بند .

(٢) مقياس الملل الأكاديمى (إعداد : الباحث)

يُعرف الملل الأكاديمى Academic Boredom بأنه شعور غير سار يظهر فى نقص الاستثارة المعرفية والاهتمام والاستمتاع بالدراسة وعدم الرضا عن مهام التعلم .

وقد تم تطوير مقياس الملل الأكاديمى بعد الاطلاع على مجموعة من الأدبيات المنشورة ومنها : (Farmer & Sundberge, 1986 ؛ فاطمة سالم العامرى ، ٢٠٠٣ ؛ ؛ يزيد عيسى السورطى ، ٢٠٠٨ ؛ تحية محمد أحمد عبد العال ، ٢٠١٢ ؛ تيسير الخوالدة ، ٢٠١٣ ؛ Dursun & Tezer, 2013 ؛ Reijseger, et al., 2013 ؛ Craparo, Faraci, Tze, Klassen, Daniels, Li, ؛ Fasciano, Carrubba & Gori, 2013 ؛ Zhang, 2013 ؛ Mercer-Lynn, Flora, Fahlman & Eastwood, ؛ مصطفى على رمضان مظلوم ، ٢٠١٤ ؛ Pekrun, Hall, Goetz & Perry, 2014) ، وتتكون الصورة المبدئية من المقياس من (٤١) بند .

(٣) استبيان الأسلوب المعرفى اللفظى–البصرى (إعداد : الباحث)

يعرف الباحث الأسلوب المعرفى الفظى–البصرى Verbalizer–Visualizer Cognitive Style بأنه يصف التفضيلات الفردية لمعالجة المعلومات البصرية مقابل المعلومات اللفظية . والأفراد ذوى الأسلوب

المعرفى البصرى يعتمدون على عمليات التصور عند تنفيذ المهام المعرفية وتكون الصور العقلية المتخيلة لدى البصريين تفصيلية وتتميز بعناصر حسية كالحركة والصوت والملمس والرائحة ويتعاملون مع كثير من المشكلات بطريقة تفكير تعتمد على التخيل المكانى ، أما الأفراد ذوى الأسلوب المعرفى اللفظى فيفضلون معالجة المعلومات بوسائل منطقية لفظية ، ويتميز اللفظيون بصور ذهنية باهتة وضعيفة ، وتركز أنماط تفكيرهم على استخدام الكلمات والتعميمات وتتسم بالتجريد .

وقد تم تطوير الاستبيان بعد الاطلاع على مجموعة من الأدبيات المنشورة منها (Warren & Good, 1979 ؛ Richardson, 1977 ؛ Parrott, ؛ Spoltore & Smock, 1983 ؛ Edwards & Wilkins, 1981 1986 ؛ Stevens, Rapp, Pfost & Johnson, 1986 ؛ Antonietti & Kozhevnikov, ؛ ١٩٩٤ ، سهير أنور محفوظ ؛ Giorgetti, 1998a,b Campos, ؛ ٢٠٠٢ ، حسام أحمد محمد ؛ Hegarty & Mayer, 2002 Lopez, Gonzalez & Amor, 2004 ؛ نصر محمود صبرى أحمد ، ٢٠٠٥؛ Massa & Mayer, 2006 ؛ نغم شريف غافل ، ٢٠١٠ ؛ Incorpora, Sala, Oliveri, Antonietti & Risoli, ؛ Kollöffel, 2012 2012 ؛ Pektas, 2014) ، وتتكون الصورة المبدئية من الاستبيان من (٢٦) بند .

ثالثاً : محكات قياس صدق المحتوى

تم فى الدراسة الحالية استخدام أربعة محكات –مستقلة– لجمع أدلة صدق المحتوى باستخدام متصل ذو مقياس رتبى رباعى النقاط Four-Point Ordinal scale وهى :

محك الارتباط Relevance : للحكم على درجة ارتباط البنود المقترحة بالمتغير المستهدف بالقياس ، ويكون المقياس الرتبى رباعى النقاط له كالتالى :

١	٢	٣	٤
غير مرتبط	مرتبط بعض الشىء	مرتبط لكنه يحتاج إلى تعديل طفيف	مرتبط بدرجة كبيرة

محك الوضوح Clarity : للحكم على درجة كون البند دقيق ومباشر وموجز ، ويكون المقياس الرتبى رباعى النقاط له كالتالى :

١	٢	٣	٤
غير واضح	واضح بعض الشىء	واضح لكنه يحتاج إلى تعديل طفيف	واضح بدرجة كبيرة

محك البساطة Simplicity : للحكم على استخدام جمل دقيقة للتعبير عن بنية البند دون زيادة أو نقصان ، ويكون المقياس الرتبى رباعى النقاط له كالتالى :

١	٢	٣	٤
غير بسيط	بسيط بعض الشىء	بسيط لكنه يحتاج إلى تعديل طفيف	بسيط بدرجة كبيرة

محك الغموض Ambiguity : للحكم على الالتباس فى معنى البند أى الحكم على احتمالات أن البند يمكن أن يُفسر (يؤول) بطرق مختلفة ، ويكون المقياس الرتبى رباعى النقاط له كالتالى :

١	٢	٣	٤
مبهم	مبهم بعض الشىء	غير مبهم لكنه يحتاج إلى تعديل طفيف	واضح المعنى بدرجة كبيرة

رابعاً : المعالجة الإحصائية للبيانات

لغرض تكمية صدق المحتوى للأدوات الثلاث تم استخدام الصيغ الإحصائية التالية :

متوسط نسبة التطابق على مستوى المقياس ACP .

مؤشر صدق المحتوى على مستوى البند CVI .

مؤشر صدق المحتوى على مستوى المقياس (S–CVI / Ave)

نسبة صدق المحتوى CVRعلى مستوى البند والمقياس.

معامل كابا k على مستوى البند والمقياس .

المؤشر r_{WG} على مستوى البند والمقياس .

نتائج اختبار فروض الدراسة وتفسيرها

نتائج اختبار الفرض الأول وتفسيرها

ينص الفرض الأول على : لا تختلف تقديرات الصيغ الإحصائية الأربع (مؤشر صدق المحتوى CVI ، معامل كابا k ، نسبة صدق المحكمين CVR ، المؤشر r_{WG}) لتكمية صدق المحتوى للمقاييس الثلاثة على مستوى البنود ، ولاختبار هذا الفرض قام الباحث بحساب أربع صيغ إحصائية (مؤشر صدق المحتوى ، نسبة صدق المحتوى ، معامل كابا ، المؤشر r_{WG}) لمحكات تقويم صدق المحتوى للأدوات الثلاث على مستوى البنود وكانت النتائج كما يلى :

(جدول (٤-١)

يتضح من الجدول (٤-١) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك الغموض على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (١ ، ٤ ، ٦ ، ٨ ، ٩ ، ١٠ ، ١٢ ، ١٣ ، ١٧ ، ١٩ ، ٢٠ ، ٢٢ ، ٢٣) لا تستبقى طبقاً لنسبة صدق المحتوى $(CVR \geq .62)$ ، والبنود (٤ ، ٨ ، ٩ ، ١٠ ، ١٣ ، ١٩ ، ٢٠ ، ٢٢ ، ٢٣) لا تستبقى طبقاً لمؤشر صدق المحتوى $(I-CVI \geq .78)$ حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (٤ ، ٩ ، ١٠ ، ١٩ ، ٢٢ ، ٢٣) لا تستبقى طبقاً لمعامل كابا $(k = .60-.74)$ حسب دراستى (1971) Fleiss ;(1984) Cicchetti .

جدول (٤-٢)

يتضح من الجدول (٤-٢) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك البساطة على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (٣ ، ٤ ، ٥ ، ٦ ، ٧ ، ٨ ، ٩ ، ١٠ ، ١٢ ، ١٤ ، ١٩ ، ٢٠ ، ٢١ ، ٢٢ ، ٢٣) لا تستبقى طبقاً لنسبة صدق المحتوى $(CVR \geq .62)$ ، والبنود (٤ ، ٦ ، ٧ ، ٨ ، ٩ ، ١٠ ، ٢٠) لا تستبقى طبقاً لمؤشر صدق المحتوى $(I-CVI \geq .78)$ حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (٤ ، ٦ ، ٨ ، ٩ ، ١٠ ، ٢٠) لا تستبقى طبقاً لمعامل كابا $(k = .60-.74)$ حسب دراستى (1971) Fleiss ;(1984) Cicchetti .

جدول (٤-٣)

يتضح من الجدول (٤-٣) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك الوضوح على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (٣ ، ٤ ، ٦ ، ١٠ ، ١٢ ، ١٣ ، ١٨ ، ١٩ ، ٢٠ ، ٢٣) لا تستبقى طبقاً لنسبة صدق المحتوى ($CVR \geq .62$) ، والبنود (٣ ، ٦ ، ٩ ، ١٠ ، ١٨ ، ١٩ ، ٢٠ ، ٢٣) لا تستبقى طبقاً لمؤشر صدق المحتوى ($I-CVI \geq .78$) حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (٩ ، ١٠ ، ١٩ ، ٢٠ ، ٢٣) لا تستبقى طبقاً لمعامل كابا ($k = .60-.74$) حسب دراستى (1971) Fleiss ;(1984) Cicchetti .

جدول (٤-٤)

يتضح من الجدول (٤-٤) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك الارتباط على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (٤ ، ١٠ ، ١٨ ، ١٩ ، ٢٠ ، ٢١ ، ٢٥) لا تستبقى طبقاً لنسبة صدق المحتوى ($CVR \geq .62$) ، والبنود (١٠ ، ١٨ ، ١٩) لا تستبقى طبقاً لمؤشر صدق المحتوى ($I-CVI \geq .78$) حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (١٠ ، ١٩) لا تستبقى طبقاً لمعامل كابا ($k = .60-.74$) حسب دراستى Cicchetti (1984); Fleiss (1971)

وبمراجعة الجداول (٤-١) إلى (٤-٤) يُستنتج أن محكات تقويم صدق المحتوى (الارتباط ، الوضوح ، البساطة ، الغموض) لا تتفق على نفس البند ، وتتفق هذه النتيجة التى تم التوصل إليها مع ما ورد فى أدبيات القياس النفسى ومنها دراسة (Ding & Hershberger, 2002: 284) .

جدول (٥-١)

يتضح من الجدول (٥-١) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك الغموض على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (١ ، ٢ ، ٣ ، ٧ ، ٢٢ ، ٢٣ ، ٢٤ ، ٢٥ ، ٢٧ ، ٢٨ ، ٢٩ ، ٣٧ ، ٣٨ ، ٣٩ ، ٤٠ ، ٤١) لا تستبقى طبقاً لنسبة صدق المحتوى $(CVR \geq .62)$ ، والبنود (٢ ، ٧ ، ٢٢ ، ٢٣ ، ٢٧ ، ٤٠ ، ٤١) لا تستبقى طبقاً لمؤشر صدق المحتوى $(I-CVI \geq .78)$ حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (٧ ، ٢٣ ، ٢٧ ، ٤٠ ، ٤١) لا تستبقى طبقاً لمعامل كابا $(k = .60-.74)$ حسب دراستى Cicchetti (1984); Fleiss (1971) .

جدول (٥-٢)

يتضح من الجدول (٥-٢) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك البساطة على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (١ ، ٢ ، ٤ ، ٧ ، ٩ ، ١٧ ، ١٨ ، ٢٣ ، ٢٤ ، ٢٦ ، ٢٧ ، ٣١ ، ٣٦ ، ٣٧ ، ٣٨ ، ٤٠ ، ٤١) لا تستبقى طبقاً لنسبة صدق المحتوى $(CVR \geq .62)$ ، والبنود (٢ ، ٤ ، ٢٤ ، ٢٦ ، ٢٧ ، ٣٦ ، ٤٠) لا تستبقى طبقاً لمؤشر صدق المحتوى $(I-CVI \geq .78)$ حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (٢ ، ٢٧ ، ٣٦ ، ٤٠) لا تستبقى طبقاً لمعامل كابا $(k = .60-.74)$ حسب دراستى Cicchetti (1984); Fleiss (1971) .

جدول (٥-٣)

يتضح من الجدول (٥-٣) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك الوضوح على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (١ ،٢ ، ٣ ، ٤ ، ٦ ، ٧ ، ١٤ ، ١٦ ، ٢٢ ، ٢٣ ، ٢٤ ، ٢٨ ، ٢٩ ، ٣٠ ، ٣٣ ، ٣٤ ، ٣٥ ، ٣٦ ، ٣٨ ، ٣٩ ، ٤٠ ، ٤١) لا تستبقى طبقاً لنسبة صدق المحتوى ($CVR \geq .62$) ، والبنود (٢ ، ٦ ، ١٦ ، ٢٢ ، ٢٣ ، ٣٥ ، ٤٠ ، ٤١) لا تستبقى طبقاً لمؤشر صدق المحتوى ($I-CVI \geq .78$) حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (٦ ، ٤١) لا تستبقى طبقاً لمعامل كابا ($k = .60-.74$) حسب دراستى Cicchetti (1984); Fleiss (1971) .

جدول (٥-٤)

يتضح من الجدول (٥-٤) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك الارتباط على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (٢ ، ٥ ، ٦ ، ١٤ ، ١٥ ، ١٦ ، ١٨ ، ٢١ ، ٢٢ ، ٢٣ ، ٢٤ ، ٢٦ ، ٢٨ ، ٢٩ ، ٣٣ ، ٣٤ ، ٣٥ ، ٣٧ ، ٣٨ ، ٣٩ ، ٤١) لا تستبقى طبقاً لنسبة صدق المحتوى ($CVR \geq .62$) ، والبنود (١٥ ، ١٦ ، ٢١ ، ٢٢ ، ٢٤ ، ٢٦ ، ٢٨ ، ٢٩ ، ٣٠ ، ٣٣ ، ٣٤ ، ٣٥ ، ٣٧ ، ٣٨ ، ٤١) لا تستبقى طبقاً لمؤشر صدق المحتوى ($I-CVI \geq .78$) حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (١٥ ، ١٦ ، ٢١ ، ٢٢ ، ٢٤ ، ٢٩ ، ٣٣ ، ٣٤ ، ٣٥ ، ٤١) لا تستبقى طبقاً لمعامل كابا ($k = .60-.74$) حسب دراستى (1971) Fleiss ; (1984) Cicchetti

جدول (٦-١)

يتضح من الجدول (٦-١) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك الغموض على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (٢ ، ٤ ، ٥ ، ٧ ، ٨ ، ٩ ، ١٠ ، ١١ ، ١٢ ، ١٣ ، ١٤ ، ١٥) لا تستبقى طبقاً لنسبة صدق المحتوى $(CVR \geq .62)$ ، والبنود (٢ ، ٥ ، ٧ ، ٨ ، ١١ ، ١٣) لا تستبقى طبقاً لمؤشر صدق المحتوى $(I-CVI \geq .78)$ حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (٢ ، ٥ ، ٨ ، ١١ ، ١٣) لا تستبقى طبقاً لمعامل كابا $(k = .60-.74)$ حسب دراستى Cicchetti (1984); Fleiss (1971) .

جدول (٦-٢)

يتضح من الجدول (٦-٢) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك البساطة على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (٢ ، ٤ ، ٥ ، ٧ ، ٨ ، ١١ ، ١٢ ، ١٣ ، ١٤ ، ١٥) لا تستبقى طبقاً لنسبة صدق المحتوى ($CVR \geq .62$) ، والبنود (٢ ، ٥ ، ٧ ، ٨ ، ١٢) لا تستبقى طبقاً لمؤشر صدق المحتوى ($I-CVI \geq .78$) حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (٢ ، ٥) لا تستبقى طبقاً لمعامل كابا ($k = .60-.74$) حسب دراستى Cicchetti (1984); Fleiss (1971) .

جدول (٦-٣)

يتضح من الجدول (٦-٣) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك الوضوح على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (٢ ، ٧ ، ٨ ، ٩ ، ١١ ، ١٢ ، ١٣ ، ١٤ ، ١٥) لا تستبقى طبقاً لنسبة صدق المحتوى ($CVR \geq .62$) ، والبنود (٢ ، ٥ ، ٧ ، ٨ ، ٩ ، ١١ ، ١٢ ، ١٣ ، ١٤ ، ١٥) لا تستبقى طبقاً لمؤشر صدق المحتوى ($I-CVI \geq .78$) حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (٢ ، ٥ ، ٩ ، ١١ ، ١٢ ، ١٣ ، ١٥) لا تستبقى طبقاً لمعامل كابا ($k = .60-.74$) حسب دراستى (1971) Fleiss ;(1984) Cicchetti .

جدول (٦-٤)

يتضح من الجدول (٦-٤) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك الارتباط على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (٢ ، ٥ ، ٦ ، ٧ ، ١١ ، ١٣) لا تستبقى طبقاً لنسبة صدق المحتوى $(CVR \geq .62)$ ، والبنود (٢ ، ٥ ، ٦ ، ٧ ، ١١ ، ١٣) لا تستبقى طبقاً لمؤشر صدق المحتوى $(I-CVI \geq .78)$ حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (٢ ، ٦ ، ٧ ، ١١ ، ١٣) لا تستبقى طبقاً لمعامل كابا $(k = .60-.74)$ حسب دراستى Cicchetti (1984); Fleiss (1971) .

وبمراجعة الجداول (٦-١) إلى (٦-٤) نجد أن محكات تقويم صدق المحتوى (الارتباط ، الوضوح ، البساطة ، الغموض) لا تتفق على نفس البند ، وتتفق هذه النتيجة التى تم التوصل إليها مع ما ورد فى أدبيات القياس النفسى ومنها دراسة (Ding & Hershberger, 2002: 284) .

نتائج اختبار الفرض الثانى وتفسيرها

ينص الفرض الثانى على : لا تختلف تقديرات الصيغ الإحصائية الخمس (مؤشر صدق المحتوى CVI ، معامل كابا k ، نسبة صدق المحكمين CVR ، متوسط نسبة التطابق ACP ، المؤشر r_{WG}) لتكمية صدق المحتوى للمقاييس الثلاث على مستوى المقياس ، ولاختبار هذا الفرض قام الباحث بحساب تقديرات خمس صيغ إحصائية (مؤشر صدق المحتوى ، متوسط نسبة التطابق ، نسبة صدق المحتوى ، معامل كابا ، المؤشر r_{WG}) لمحكات تقويم صدق المحتوى للأدوات الثلاث (مقياس الهزيمة النفسية ، مقياس الملل الأكاديمى ، استبيان الأسلوب المعرفى) على مستوى المقياس وكانت النتائج كما يلى :

جدول (٧)

الملخص

هدفت هذه الدراسة إلى استخدام خمس صيغ إحصائية لتكمية صدق المحتوى هى : نسبة صدق المحتوى CVR ، معامل كابا k ، متوسط نسبة التطابق ACP ، نسبة صدق المحتوى CVI ، والمؤشر r_{WG} لثلاثة اختبارات نفسية (مقياس الهزيمة النفسية ، مقياس الملل الأكاديمى ، استبيان الأسلوب المعرفى البصرى–اللفظى)وجميعها من إعداد الباحث ، وبلغ حجم عينة المحكمين (١٠) محكمين متخصصين .

وتوصلت نتائج الدراسة إلى :

يُعطى المؤشر r_{WG} أقل تقدير لمحكات صدق المحتوى على مستوى البنود Item-Level وتليه نسبة صدق المحتوى CVR ويليها معامل كابا k وأخيراً مؤشر صدق المحتوى CVI كأكبر تقدير لكافة المقاييس المستخدمة .

يُعطى المؤشر r_{WG} أقل تقدير لمحكات صدق المحتوى على مستوى المقياس Scale–Level وتليه نسبة صدق المحتوى CVR ثم معامل كابا k ويليه مؤشر صدق المحتوى CVI وأخيراً متوسط نسبة التطابق ACP كأكبر تقدير لكافة المقاييس المستخدمة .

لا تتفق تقديرات محكات تقويم صدق المحتوى (الارتباط ، الوضوح ، البساطة ، الغموض) على نفس أداة القياس على مستوى البنود وعلى مستوى المقياس .

وإجمالاً يرى الباحث أن الطرق الكمية التى استخدمت للتحقق من صدق المحتوى لأدوات القياس الثلاث تزيد مقدار المعلومات المتاحة لتحرى الخصائص السيكومترية ، كما تزيد الثقة فى صدق محتوى هذه الأدوات الجديدة .

تكمية صدق المحتوى
فى بعض الاختبارات النفسية

يتضح من الجدول (٦-٤) السابق أن المؤشر r_{WG} يعطى أقل تقدير لمحك الارتباط على مستوى البنود وتليه نسبة صدق المحتوى ويليها معامل كابا وأخيراً تقدير مؤشر صدق المحتوى كأعلى تقدير ، وهذه النتيجة لا تتفق مع ما توصلت إليه نتائج دراسة (Bayrak & Yurdugul, 2012) بوجود اتساق فى نتائج مؤشر صدق المحتوى ونسبة صدق المحتوى ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

ومن الجدول أيضاً يتضح أن البنود (٢ ، ٥ ، ٦ ، ٧ ، ١١ ، ١٣) لا تستبقى طبقاً لنسبة صدق المحتوى $(CVR \geq .62)$ ، والبنود (٢ ، ٥ ، ٦ ، ٧ ، ١١ ، ١٣) لا تستبقى طبقاً لمؤشر صدق المحتوى $(I-CVI \geq .78)$ حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، والبنود (٢ ، ٦ ، ٧ ، ١١ ، ١٣) لا تستبقى طبقاً لمعامل كابا $(k = .60-.74)$ حسب دراستى Cicchetti (1984); Fleiss (1971) .

وبمراجعة الجداول (٦-١) إلى (٦-٤) نجد أن محكات تقويم صدق المحتوى (الارتباط ، الوضوح ، البساطة ، الغموض) لا تتفق على نفس البند ، وتتفق هذه النتيجة التى تم التوصل إليها مع ما ورد فى أدبيات القياس النفسى ومنها دراسة (Ding & Hershberger, 2002: 284) .

نتائج اختبار الفرض الثانى وتفسيرها

ينص الفرض الثانى على : لا تختلف تقديرات الصيغ الإحصائية الخمس (مؤشر صدق المحتوى CVI ، معامل كابا k ، نسبة صدق المحكمين CVR ، متوسط نسبة التطابق ACP ، المؤشر r_{WG}) لتكمية صدق المحتوى للمقاييس الثلاث على مستوى المقياس ، ولاختبار هذا الفرض قام الباحث بحساب تقديرات خمس صيغ إحصائية (مؤشر صدق المحتوى ، متوسط نسبة التطابق ، نسبة صدق المحتوى ، معامل كابا ، المؤشر r_{WG}) لمحكات تقويم صدق المحتوى للأدوات الثلاث (مقياس الهزيمة النفسية ، مقياس الملل الأكاديمى ، استبيان الأسلوب المعرفى) على مستوى المقياس وكانت النتائج كما يلى :

يتضح من الجدول (٧) السابق استقرار واتساق أداء الصيغ الإحصائية الخمس المستخدمة لتكمية صدق المحتوى على مستوى المقياس عبر أدوات القياس الثلاث حيث تساوت قيمتى مؤشر صدق المحتوى ومتوسط نسبة التطابق وكانا أعلى قيمة من بين قيم الصيغ الإحصائية الخمس وتليهما قيمة معامل كابا ثم قيمة نسبة صدق المحتوى وأخيراً المؤشر r_{WG}كأقل تقدير من بين تقديرات الصيغ الخمس ، وتتفق هذه النتيجة جزئياً مع نتيجة دراسة (Lindell & Brandt, 1999) بأن تقدير المؤشر r_{WG} أقل من تقدير نسبة صدق المحتوى ، كما تتفق النتيجة تماماً مع نتيجة دراسة (Beckstead, 2009: 1274) أن مؤشر صدق المحتوى يوفر تقديراً متضخماً Inflated Estimate لصدق المحتوى مقارنة بالصيغ الإحصائية الأخرى .

كما يتضح من الجدول السابق أيضاً أن مقياس الهزيمة النفسية استوفى فقط محك الارتباط طبقاً لمتوسط نسبة التطابق $(ACP \geq .9)$

حسب توصية دراسة (Waltz, Strickland & Lenz, 2005: 178) ، واستوفى فقط محك الارتباط طبقاً لمؤشر صدق المحتوى (S–*CVI* / Ave ≥.9) حسب توصية دراسة (Jingcheng, Xiankun & Zhenqiu, 2012) ، واستوفى المحكات الأربعة طبقاً لنسبة صدق المحتوى (*CVR* ≥.62) ، واستوفى المحكات الأربعة طبقاً لمعامل كابا (*k* =.60–.74) حسب دراستى Cicchetti (1984); Fleiss (1971) .

وأن مقياس الملل الأكاديمى لم يستوف المحكات الأربعة طبقاً لمتوسط نسبة التطابق (*ACP* ≥.9) ، ولم يستوف المحكات الأربعة طبقاً لمؤشر صدق المحتوى (S–*CVI* / Ave ≥.9)، واستوفى ثلاثة محكات (الغموض ، البساطة ، الوضوح) طبقاً لنسبة صدق المحتوى (*CVR* ≥.62) ، واستوفى المحكات الأربعة طبقاً لمعامل كابا (*k* =.60–.74) حسب دراستى Cicchetti (1984); Fleiss (1971)

وأن استبيان الأسلوب المعرفى لم يستوف جميع المحكات طبقاً لمتوسط نسبة التطابق (*ACP* ≥.9) ومؤشر صدق المحتوى (S–*CVI* / Ave ≥.9) ، واستوفى المحكات الأربعة طبقاً لنسبة صدق المحتوى (*CVR* ≥.62) ومعامل كابا (*k* =.60–.74) حسب دراستى Cicchetti (1984); Fleiss (1971) .

وهذا يعنى أن تقديرات محكات تقويم صدق المحتوى (الارتباط ، الوضوح ، البساطة ، الغموض) لا تتفق على نفس أداة القياس ، وتتفق هذه النتيجة التى تم التوصل إليها مع ما ورد فى أدبيات القياس النفسى ومنها دراسة (Ding & Hershberger, 2002: 284) بأنه ربما لا تتفق محكات تقويم صدق المحتوى لنفس أداة القياس

مستخلص الدراسة

بمراجعة نتائج اختبار فرضى الدراسة يمكن استخلاص الآتى :

يتوقع ارتفاع تقديرات الصيغ الإحصائية المستخدمة لتكمية صدق المحتوى على مستوى المقياس بعد حذف/تعديل البنود غير الملائمة من حيث محكات الغموض والبساطة والوضوح والارتباط .

توجد صعوبة فى استخدام أربعة محكات لتقويم صدق المحتوى ، فقد أفاد المحكمون بصعوبة التمييز بين محكات الوضوح والغموض والبساطة

يتوقع أن تقل مشكلات استخلاص وتسمية العوامل فى التحليل العاملى الاستكشافى إذا أجرى قبلها تكمية لصدق المحتوى على مستوى البنود مصحوب بتنفيذ تعديلات المحكمين على وظائف البنود المحددة طبقاً لمحكات التقويم .

لا تتفق محكات تقييم صدق المحتوى على نفس البند .

توصيات الدراسة

فى ضوء الإطار النظرى والنتائج التى تم التوصل إليها فى الدراسة الحالية يوصى الباحث الحالى بالتالى :

وضع معايير استرشادية لاختيار المحكمين تشمل (التخصص الدقيق ، الخبرة ، الدرجة العلمية).

ارفاق التعريف الإجرائى للمتغير المستهدف بالقياس ضمن الأداة المقدمة للتحكيم .

من المهم أن يكون قرار المحكم إما ثنائى أو رباعى التدرج يختزل إلى ثنائى أثناء تحليل استجابات المحكمين باستخدام صيغ تكمية صدق المحتوى ، ولا يفضل متصل القرارات الثلاثى أو الخماسى للحكم على قيام البنود بوظيفتها من حيث الارتباط والوضوح والغموض والبساطة.

استخدام معامل كابا لعدم وجود تقديرات صفرية أو سالبة أو تقديرات متضخمة مما يجعله الصيغة الإحصائية الأسهل فى التفسير .

الحذر عند استخدام نسبة صدق المحتوى دون الوعى بدلالة التقديرات الصفرية والسالبة والتى تعنى أن الاتفاق بين المحكمين لا يرقى للاعتماد عليه لاستبقاء البند .

ضرورة فحص صدق المحتوى على مستوى البنود وعلى مستوى المقياس معاً .

الخطوة التالية لتكمية صدق المحتوى تتطلب التحليل الكيفى لتعليقات المحكمين لتعديل أو تنقيح البنود المتبقية قبل أى تحليلات لاحقة باعتبار صدق المحتوى متطلباً قبلياً لصدق البناء والصدق المرتبط بالمحك .

المصادر والمراجع

١. إقبال زين العابدين درندرى (٢٠١٢) . اختبارات القبول بالجامعات السعودية فى ضوء التوجهات الحديثة فى صدق الاختبار . **المؤتمر الدولى الأول للقياس والتقويم "معايير القبول فى التعليم العالى".**

المركز الوطنى للقياس والتقويم فى التعليم العالى. المملكة العربية السعودية .

٢. أمين على محمد سليمان (٢٠١٠) . **القياس والتقويم فى العلوم الإنسانية أسسه وأدواته وتطبيقاته** . القاهرة : دار الكتاب الحديث.

٣. تحية محمد أحمد عبد العال (٢٠١٢) . الضجر وعلاقته بمعنى الحياة لدى طلاب الجامعة (دراسة فى سيكولوجية المضجر) . **مجلة كلية التربية . جامعة بنها** . ٢٣ (٩٢) ، ٤٣٣ – ٥٢١ .

٤. تيسير الخوالدة (٢٠١٣) . الملل الأكاديمى لدى طلبة كلية العلوم التربوية فى جامعة آل البيت . **المنارة** . ١٩(١) ، ٧٩ – ٩٧ .

٥. جاى ل. ر. ، ميلز ج. ، إيراسيان ب. (٢٠١٢) . **البحث التربوى كفايات للتحليل والتطبيقات** . ترجمة : صلاح الدين محمود علام . عمان : دار الفكر .

٦. حسام أحمد محمد (٢٠٠٢) . سيكولوجية الصور العقلية . **مجلة التربية** . ٣١(١٤٢) ، ١٧٢ – ١٩٠ .

٧. حيدر إبراهيم ظاظا (٢٠١١) . درجة توافق دلالات صدق وثبات الاختبارات المقننة المستخدمة فى **رسائل الماجستير** المقدمة فى كليات التربية فى الجامعات الأردنية مع دلالات صورها الأصلية . **دراسات : العلوم التربوية** . ٣٨(٢) ، ٢٤٠٠ – ٢٤١٦ .

٨. رجاء محمود أبو علام (٢٠٠٧) . **مناهج البحث فى العلوم النفسية والتربوية** . القاهرة : دار النشر للجامعات .

٩. رمزية الغريب (١٩٩٦) . **التقويم والقياس النفسى والتربوى** . القاهرة : مكتبة الأنجلو المصرية .

١٠. سامية سمير شحاتة (٢٠٠٣) . مدى فعالية صدق المحكمين بالمقارنة بأنواع الصدق الأخرى : **دراسة تحليلية** . **رسالة دكتوراة غير منشورة** . كلية الآداب . جامعة المنيا .

١١. سهير أنور محفوظ (١٩٩٤) . التخيل العقلى لدى طالبات الجامعة فى علاقته بالأسلوب المعرفى . **المجلة المصرية للدراسات النفسية** . (٨) ، ١٦٩ – ١٩٥ .

١٢. صالح سليمان البقعاوى (٢٠١٤) . بعض أخلاقيات محكمى البحث العلمى من منظور التربية الإسلامية . **مجلة التربية للبحوث التربوية والنفسية والاجتماعية** . كلية التربية . جامعة الأزهر . العدد (١٥٩) – الجزء الثانى ، ١٩٥ – ٢٠٩ .

١٣. صفوت فرج (١٩٩٧) . **القياس النفسى** . الطبعة الثالثة . القاهرة : مكتبة الأنجلو المصرية .

١٤. عبد الله على القاطعى (١٩٩٣) . علاقة صدق المحكمين بالمفهوم الإحصائى لصدق البنود . **دراسات نفسية** . ٣(١) ، ٥٣ – ٦١.

١٥. عبد الفتاح إبراهيم القرشى (١٩٩٧) . مقياس السلوك الدراسى لطلبة الجامعة . **المجلة التربوية** . ١١(٤٣) ، ١١١ – ١٥١ .

١٦. عزت عبد الحميد (٢٠١١) . **الإحصاء النفسى والتربوى : تطبيقات باستخدام برنامج SPSS 18** . القاهرة : دار الفكر العربى .

١٧. فاروق السيد عثمان ، عبد الهادى السيد عبده (١٩٩٥) . **الإحصاء التربوى والقياس النفسى** . القاهرة : دار المعارف .

١٨. فاروق عبد الفتاح موسى (١٩٩٠) . **القياس النفسى والتربوى للأسوياء والمعاقين** . القاهرة : مكتبة النهضة المصرية .

١٩. فاطمة سالم العامرى (٢٠٠٣) . المشكلات الأكاديمية لدى طلبة جامعة الإمارات العربية المتحدة .**مجلة كلية التربية . جامعة الإمارات العربية المتحدة** . العدد (٢٠) ، ١١٩ –١٨٣ .

٢٠. فؤاد البهى السيد (١٩٧٨) . **علم النفس الإحصائى وقياس العقل البشرى** . الطبعة الثالثة . القاهرة : دار الفكر العربى .

٢١. فضل إبراهيم عبد الصمد (٢٠١٣) . الهزيمة المصرية (مقياس الهزيمة النفسية فى البيئة المصرية) . **مجلة البحث فى التربية وعلم النفس** . كلية التربية . جامعة المنيا . ٢٦(١) ، – .

٢٢. محمد أحمد الرفوع ، تيسير خليل القيسى ، أحمد عودة القرارعة (٢٠٠٩) . علاقة الكفاءة الذاتية المدركة بالقدرة على حل المشكلات لدى طلبة جامعة الطفيلة التقنية فى الأردن . **المجلة التربوية** . ٢٣(٩٢) ، ١٨١ – ٢١٤ .

٢٣. محمد إسماعيل عمران ، عبد الله السيد عبد الجواد ، فهد عبد الله الدليم (١٩٩٤) . **مبادىء القياس والتقويم فى البيئة الإسلامية** .

٢٤. محمد السعيد عبد الجواد أبو حلاوة (٢٠١٢) . الهزيمة النفسية : ماهيتها ، مؤشراتها ، محدداتها ، تداعياتها ، والوقاية منها "دراسة فى بناء المفهوم" . **مجلة الدراسات التربوية والإنسانية** . كلية التربية . جامعة دمنهور . ٤ (٣) ، ١٧٧ – ٢٦٢ .

٢٥. محمد السعيد عبد الجواد أبو حلاوة ، راشد مرزوق راشد (٢٠١٣). البنية العاملية والتحليل التمييزى للهزيمة النفسية فى ضوء بعض المتغيرات النفسية لدى طلاب الجامعة (نموذج مقترح) . **دراسات عربية فى التربية وعلم النفس** . ع (٣٧) . الجزء (٣)

٢٦. مصطفى على رمضان مظلوم (٢٠١٤) . فعالية برنامج ارشادى لتخفيف الضجر لدى عينة من طلاب الجامعة . **دراسات عربية فى التربية وعلم النفس** . العدد (٥٢) ، الجزء الأول ، .

٢٧. ميخائيل أسعد (١٩٩٠) . **الإحصاء النفسى وقياس القدرات الإنسانية** . بيروت : دار الآفاق الجديدة .

٢٨. نصر محمود صبرى أحمد (٢٠٠٥) . البناء العقلى للتخيل العقلى فى علاقته بالابتكارية وحل المشكلات . **رسالة دكتوراة غير منشورة** . كلية التربية . جامعة الزقازيق .

٢٩. نغم شريف غافل (٢٠١٠) . الأسلوب المعرفى (اللفظى–البصرى) وعلاقته بالتفضيل الجمالى لدى طلبة الجامعة . **رسالة ماجستير غير منشورة** . كلية الآداب . جامعة بغداد .

٣٠. يزيد عيسى السورطى (٢٠٠٨) . درجة تعرض طلبة تخصص معلم صف فى الجامعة الهاشمية للملل الأكاديمى وعلاقتها ببعض المتغيرات . دراسات ، **العلوم التربوية** . ٣٥(١) ، ٥٧– ٦٦ .

31. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). Standards **for educational and psychological testing.** Washington, D.C.: American Educational Research Association.

32. American Psychological Association (2010). **Publication manual of the American Psychological Association (6th ed.)**. Washington, DC: Author.

33. Antonietti, A., & Giorgetti, M. (1998a). A study of some psychometric properties of the Verbalizer-

Visualizer questionnaire. **Journal of Mental Imagery**. 20(3&4), 59-68.

34. Antonietti, A., & Giorgetti, M. (1998b). **The Verbalizer-Visualizer questionnaire**: A review. Perceptual and Motor Skills. 86(), 227-239.
35. Ayre, C., & Scally, A. J. (2014). C**ritical values for Lawshe's content validity ratio**: Revisiting the original methods of calculation . Measurement and Evaluation in Counseling and Development. 47(1), 79 –86.
36. Bayrak, F., & Yurdugul, H. (2012). Content validity measures in scale development studies: comparison of content validity index and Kappa statistics. H. U. **Journal** of Education. 2, 246-271.
37. Beckstead, J. W. (2009). Content validity is naught. International **Journal** of Nursing Studies. 46, 1274-1283.
38. Campos, A., Lopez, A., Gonzalez, M. A., & Amor, A. (2004). **Imagery factors in the Spanish version of the verbalizer-visualizer questionnaire.** Psychological Reports. 94(), 1149-1154.
39. Carretero-Dios, H., & Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. International **Journal** of Clinical and Health Psychology, 7, 863-882.
40. Cicchetti, D. V. (1984). **On a model for assessing the security of infantile attachment**: Issues of observer reliability and validity. Behavioral and Brain Sciences, 7, 149-150.
41. Cohen, J. (1960). A coefficient of agreement **for nominal scales**. Educational and Psychological Measurement. 20, 37-46.
42. Craparo, G., Faraci, P., Fasciano, S., Carrubba, S., & Gori, A. (2013). **A factor analytic study of the**

boredom proneness scale (BPS). Clinical Neuropsychiatry. 10(3,4). 164-170.

43. Davis L.L. (1992). **Instrument review**: Getting the most from a panel of experts. Applied Nursing Research. 5, 194-197.
44. Delgado-Rico, E., Carretero-Dios, H., & Ruch, W. (2012). The use of expert judges in scale development: Implications for improving face validity of measures of unobservable constructs. International **Journal** of Clinical and Health Psychology. 12(3), 449-460.
45. DeVellis, R. F. (2003). Scale development: **Theory and applications**. (2nd ed.). Thousand Oaks, CA: Sage Publications.
46. Ding, C. S. & Hershberger, S. L. (2002). **Assessing content validity and content equivalence using structural equation modeling. Structural Equation Modeling**. 9(2), 283-297.
47. Downey, M., Collins, M., & Browning, W. (2002). Predictors of success in dental hygiene education: a six-year review. **Journal** of Dental Education, 66 (11), 1269-1273.
48. Dursun, P., & Tezer, E. (2013). **Turkish adaptation of the boredom proneness scale** short-form. Procedia-Social and Behavioral Sciences. (84), 1550-1554.
49. Ebn Ahmady, A., Yazdani, S., Valian, A., Amiri, Z., Mortazavi, F., & Lando, H. A. (2013). **Psychometric properties of a Farsi version of the authentic assessment perception questionnaire**. **OHDM. 12(4), 284-288.**

50. Farahani, M. A., Ghasemi, H. S. E., Nikpeyma, N., Fereidouni, Z., & Rassouli, M. (2015) Development and psychometric evaluation of the nursing instructors' clinical teaching performance inventory. Global **Journal** of Health Science. 7(3), 30-36.
51. Farmer, R. F., & Sundberge, N. D. (1986). Boredom proneness: The development and correlates of a new scale. **Journal** of Personality Assessment. 50, 4-17.
52. Fleiss, J. (1971). **Measuring nominal scale agreement among many raters**. Psychological Bulletin, 76, 378-382.
53. Fung, K. (2010). The **initial development and content validity of an Asperger's** Syndrome Self-Screening Instrument for adults (Unpublished Master Thesis). University of Saskatchewan.
54. Gelfand, D. M., & Hartmann, D. P. (1975). **Child behavior analysis and therapy**. New York: Pergamon.
55. Gholami, M., Pakdaman, A., Jafari, A., & Virtanen, J. I. (2014). Knowledge of and attitudes towards periodontal health among adults in Tehran. Eastern Mediterranean Health **Journal**. 20(3), 196-202.
56. Gilbert, P. & Allen, S. (1998). The role of defeat and entrapment (arrested flight) in depression: An exploration of an evolutionary view. **Psychological** Medicine. 28, 585-598.
57. Gilbert, P. (2001a). Evolutionary approaches to psychopathology: The role of natural defenses. Australian and New Zealand **Journal** of Psychiatry. 35, 17-27.
58. Gilbert, P. (2001b). Depression and stress: A biopsychosocial exploration of evolved functions and mechanisms Stress: The International **Journal** of the Biology of Stress. 4, 121-135.
59. Gilbert, P., Allan, S., Brough, S., Melley, S. & Miles, J. (2002). Anhedonia and positive affect:

relationship to social rank, defeat and entrapment. **Journal** of Affective Disorders. 71, 141-151.

60. Gilbert. P. (2007). **Psychotherapy and Counseling for depression. London**: Sage.
61. Grant, J.S., & Davis, L.T. (1997). **Selection and use of content experts in instrument development. Research in Nursing & Health**, 20, 269–274.
62. Haynes, S.N., Richard, D.C.S., & Kubany, E.S. (1995). **Content validity in psychological assessment**: A functional approach to concepts and methods. Psychological Assessment. 7(3), 238-247.
63. Incorpora, C., Sala, G., Oliveri, S., Antonietti, A., & Risoli, A. (2012). **Verbalizer-visualizer style in brain-lesioned patients**: Does rehabilitation matter?. Review of Psychology Frontier. 1(1), 33-37.
64. James, L. R., Demaree, R. G. & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. **Journal** of Applied Psychology. 69, 85-98.
65. James, L. R., Demaree, R. G. & Wolf, G. (1993). r_{WG}: An assessment of within-group interrater agreement. **Journal** of Applied Psychology. 73, 306-309.
66. Janda, L. H. (1998). Psychological Testing: **Theory and Application. Boston**: Allyn and Bacon.
67. Jingcheng, S., Xiankun, M., & Zhenqiu, S. (2012). Content validity index in scale development. **Journal** of Central South University (Medical Sciences). 37(2), 152-155.
68. Kollöffel, B. (2012). Exploring **the relation between visualizer-verbalizer cognitive styles and**

performance with visual or verbal learning material. Computer & Education. 58(2), 697-706.

69. Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). **Revising the Visualizer-Verblizer dimension**: Evidence for two types of visualizers. Cognition and Instruction. 20(1), 47-77.
70. Kubiszyn, T., & Borich, G. (1990). **Educational testing and measurement**: Classroom application and practice. (3rd ed.). London: Scott, Foresman and Company.
71. Landis, J. R., & Koch, G. G. (1977). **The measurement of observer agreement for categorical data. Biometrics**, 33, 1159-1174.
72. Lawlis, G. F., & Lu, E. (1972). J**udgment of counseling process**: reliability, agreement, and error. Psychological Bulletin. 78, 17-20.
73. Lawshe, C.H. (1975). **A quantitative approach to content validity**. Personnel Psychology, 28, 563–575.
74. Li, H.C.W., & Lopez, V. (2004). **Psychometric evaluation of the Chinese version of the State Anxiety Scale for Children**. Research in Nursing & Health, 27, 198– 207.
75. Lindell, M.K., & Brandt, C.J. (1997). **Measuring interrater agreement for ratings of a single target. Applied Psychological Measurement**. 21(3), 271-278.
76. Lindell, M.K., & Brandt, C.J. (1999). **Assessing interrater agreement on the job relevance of a test**: A comparison of the CVI, T, $r_{WG(J)}$, and $r^{*}_{WG(J)}$ indexes**. Journal of Applied Psychology**, 84, 640– 647.

77. Lindgren, K. (2005). Testing the Health Practices in Pregnancy Questionnaire-II. **Journal** of Obstetric, Gynecologic, & Neonatal Nursing, 34, 465– 472.
78. Lotfi, R., Tehrani, R. T., Yaghmaie, F., & Hajizadeh, E. (2014). Developing a valid and reliable instrument to predict the protective sexual behaviors in women at risk of human immunodeficiency virus. Iranian Red Crescent Medical **Journal** . 16(9), 1-7.
79. Lynn, M. R. (1986). **Determination and quantification of content validity**. Nursing Research. 35, 382–385.
80. Martuza, V.R. (1977). **Applying norm-referenced and criterion-referenced measurement in education. Boston**: Allyn & Bacon.
81. Massa, L. J., & Mayer, R. E. (2006). **Testing the ATI hypothesis**: Should multimedia instruction accommodate verbalizer-visualizer cognitive style?. Learning and Individual Differences. 16, 321-335.
82. Mastaglia, B., Toye, C., & Kristjanson, L.J. (2003). En**suring content validity in instrument development**: Challenges and innovative approaches. Contemporary Nurse, 14, 281–291.
83. McCain, N. L. (1984). A test of Cohen's development model for professional socialization with baccalaureate nursing students. Unpublished **doctoral dissertation**, the university of Alabama in Birmingham.
84. Mercer-Lynn, K. B., Flora, D. B., Fahlman, S. A., & Eastwood, J. D. (2013). **The measurement of boredom:** differences between existing self-report scales. Assessment. 20(5), 585-596.
85. Messick, S. (1989). Validity. In R. L. Linn (ed.). Educational measurement (PP. 13-103). (3rd ed.). New York: **MacMillan.**

86. Nordin, S., Elf, M., McKee, K., & Wijk, H. (2015). **Assessing the physical environment of older people's residential care facilities**: development of the Swedish version of the Sheffield Care Environment Assessment Matrix (S-SCEAM). BMC Geriatr. 15(1),
87. Pekrun, R., Hall., N. C., Goetz, T., & Perry, R. P. (2014). Boredom and academic achievement: Testing a model of reciprocal causation. **Journal of Educational Psychology**. 106(3), 696-710.
88. Pektas, S. T. (2014). **Correlations between the visualizer/imager cognitive style and achievement in digital modeling tasks**. Procedia-Social and Behavioral Sciences. 116, 5053-5057.
89. Polit, D.F., & Beck, C.T. (2006). **The content validity index**: are you sure you know what's being reported? Critique and recommendations. Research in Nursing & Health, 29, 489- 497.
90. Polit, D.F., Beck, C.T., & Owen, S. (2007). Is the CVI an acceptable indicator of content validity? **Research in Nursing & Health**, 30, 459-467.
91. Popham, W. J. (1990). **Modern educational measurement**: A practitioner's perspective. . (2nd ed.). Englewood Cliffs, New Jersey: Prentice Hall, Inc.
92. Reijseger, G., Schaufeli, W. B., Peeters, M. C. W., Taris, T. W., van Beek, I., & Ouweneel, E. (2013). **Anxiety, Stress, & Coping**. 26(5), 508-525.
93. Richardson, A. (1977) Verbalizer-Visualizer: a cognitive style dimension. **Journal** of Mental Imagery, 1, 109-126.
94. Sharp, M., & Rhia, M. B. A, (2010). **Development of an instrument to measure students' perceptions of information technology fluency skills**: Establishing content validity. Perspectives in Health Information Management. 1-10.

95. Silver, B. B., Smith, E. V., Jr., Greene, B. A. (2001). A study strategies self-efficacy instrument for use with community college students. Educational and Psychological Measurement. 61(5), 849-865.
96. Sireci, S. G. (1998). **Evaluating content validity using multidimensional scaling**. Paper presented at the annual meeting of the AERA (San Diego, CA., April).
97. Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. **Psicothema.** 26(1), 100-107.
98. Slavin, R. E. (1997). **Educational Psychology: Theory and Practice**. (5th ed.). Boston: Allyn and Bacon.
99. Spoltore, J. D., & Smock, D. J. (1983). **The Verbalizer-Visualizer questionnaire: Additional normative data**. Perceptual and Motor Skills. (56), 382.
100. Stevens, M. J., Rapp, B. J., Pfost, K. S., & Johnson, J. J. (1986) **Further evidence of the stability of the Verbalizer-Visualizer Questionnaire**. Perceptual and Motor Skills. 62, 301-302.
101. Suen, H. K., & Ary, D. (2014). **Analyzing quantitative behavioral observation data**. Hillsdale, NJ: Psychology Press.
102. Todman, M. (2013). The dimensions of state boredom: Frequency, duration, unpleasantness, consequences and causal attributions. **Educational Research International**. 1(1), 32-40.
103. Treloar, H., Martens, M. P., & McCarthy, D. M. (2015). **The protective behavioral strategies scale-20: Improved content validity of the**

serious harm reduction subscale. Psychological Assessment.

104. Tze, R. M., Klassen, R. M., Daniels, L. M., Li, J. C.-H., & Zhang, X. (2013). A cross-cultural validation of the learning-related boredom scale (LRBS) with Canadian and Chinese college students. **Journal of Psychoeducational Assessment**. 31(1), 29-40.
105. Vadlin, S., Aslund, C., Nilsson, K. W. (2015). Development and content validity of a screening instrument for gaming addiction in adolescents: The gaming addiction identification test (GAIT). **Scandinavian Journal of Psychology**.
106. van Schelven, A. R., Dikken, J., Sillekens, L. G. M., Oldenhuis, D. D., Schuurmans, M. J., & Hoogerduijn, J. G. (2015). Content validation of the Dutch version of the "older patient in acute care survey", an instrument to measure the attitude of hospital nurses towards older patients. **International Journal of Clinical Medicine. 6, 7-18**.
107. Waltz, C.F., & Bausell, R.B. (1981). **Nursing research: Design, statistics, and computer analysis**. Philadelphia: F. A. Davis.
108. Waltz, C.F., Strickland, O.L., & Lenz, E.R. (2005). **Measurement in nursing and health research (3rd ed.)**. New York: Springer Publishing Co.
109. Wilson, F.R., Pan, W., & Schumsky, D.A. (2012). **Recalculation of the critical values for Lawshe's content validity ratio**. Measurement and Evaluation in Counseling and Development, 45(3), 197-210.
110. Woolfolk, A. (2005). **Educational Psychology**. Boston: Allyn and Bacon.

111. Wynd, C. A., & Atkins Schaefer, M. (2002). **The Osteoporosis Risk Assessment Tool: Establishing content validity through a panel of experts**. Applied Nursing Research, 16, 184-188.
112. Wynd, C.A., Schmidt, B., & Schaefer, M.A. (2003). Two quantitative approaches for estimating instrument content validity. Western **Journal of Nursing Research**. 25, 508-518.
113. Yaghmaei F, Mohammadi, S. H., & Majd, A.H. (2013). Developing and measuring psychometric properties of "quality of life questionnaire in infertile couples. The International **Journal of Community Based Nursing and Midwifery**. 1(4), 238-245.
114. Yaghmaie, F. (2003). Content validity and its estimation. **Journal of Medical Education**. 3(1), 25-27.
115. Yuen, E. Y. N., Knight, T., Dodson, S., Ricciardelli, L., Susan Burney, S., & Livingston, P. M. (2014). **Development of the Health literacy of caregivers scale - cancer (HLCS-C): item generation and content validity testing**. BMC Family Practice
116. Zamanzadeh, V., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., Nikanfar, A.-R., & Ghahramanian, A. (2014). **Details of content validity and Objectifying it in instrument development**. Nursing Practice Today. 1(3), 163-171.
117. Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning

(DIF): logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores. Ottawa ON: **directorate of human resources research and evaluation**, department of national defense.

Quantification of Content Validity in some Psychological Tests

This study aimed to use five statistical forms to quantify content validity (content validity ratio CVR, Kappa coefficient k, average congruency percentage ACP, content validity index CVI, and r_{WG} index) for

three psychological intruments (mental defeat scale, boredom scale, and verblizer-visulizer cognive style questionnaire). A panel of (10) content experts were asked to rate each instrument item in terms of relevance, clarity, simplicity, and ambiguity on a four-point ordinal scale.

The analysis revealed that:

The r_{WG} index produce the minimum rate for the content validy criteria on the item-level followed by content validity ratio CVR and then Kappa coefficient (k) and finally content validity index CVI as the maximum rate for all the instruments used.

The r_{WG} index produce the minimum rate for the content validy criteria on the scale-level followed by content validity ratio CVR followed by Kappa coefficient (k) and then content validity index CVI and finally average congruency percentage ACP as the maximum rate for all the instruments used.

There was inconsistency of content validity ratio CVR, Kappa coefficient k, average congruency percentage ACP, content validity index CVI, and rWG index for any instrument on the item-level and scale level.

Generaly, quantitative methods used to confirm the content validity of the three new tools increase the amount of information available for detecting the psychometric properties and increase confidence in the content validity of the tools.

ملحق (١) بنود مقياس الهزيمة النفسية

م	البنود
١	أشعر بأننى ليس لى بصمة فى الحياة
٢*	أشعر بأننى شخص ناجح
٣	أشعر بأننى منهزم من الحياة
٤*	أشعر بأننى فى الأساس فائزاً

٥	أشعر بأننى فقدت مكانتى فى الحياة
٦	أشعر بأن الحياة تعاملنى كشىء مهمل
٧	أشعر بالعجز
٨	أشعر بأن ثقتى قد ولت عنى
٩*	أشعر بالقدرة على التعامل مع ما تلقيه الحياة على وجهى
١٠	أشعر بأننى غرقت أسفل السلم
١١	أشعر تماماً بعدم قدرتى على العمل
١٢	أشعر بأننى أحد الخاسرين فى الحياة
١٣	أشعر بأننى استسلمت
١٤	أشعر بالتراجع والدونية
١٥	أشعر بأننى خسرت معارك هامة فى الحياة
١٦	أشعر بأنه لا حول ولا قوة لى
١٧	أشعر بأننى شخص مدمر
١٨	أشعر بأننى أريد أن أموت
١٩	لا أهتم بما حدث لى بعد الآن
٢٠	أشعر تماماً بأننى تحت رحمة ما كان يحدث لى
٢١	أشعر بالذل وأننى أفقد الشعور بكرامتى
٢٢	أشعر بأننى أفقد مقاومتى الداخلية
٢٣	أشعر بأننى أقل من إنسان
٢٤	أشعر أنى تافه
٢٥	أشعر بأنى لا ماض ولا مستقبل لى
٢٦	أشعر أنى غير قادر على تطوير ذاتى
٢٧	أشعر بعدم الدافعية للحياة

ملحق (٢) بنود مقياس الملل الأكاديمى

م	البنود
١	الدراسة مملة ورتيبة
٢	يتوه عقلى حين أستذكر
٣	عندما أدرس هذه المقررات المملة ، أقضى وقتى مفكراً فى الزمن المتبقى
٤	استراتيجيات التدريس تجعلنى أحس بأن وقت المحاضرات يمر ببطء شديد
٥	يغلب على محتوى المناهج الدراسية الحشو والتكرار أو التداخل
٦	محتوى معظم المناهج الدراسية لا يتناسب مع ميولى واتجاهاتى
٧	أود تأجيل الأعمال المملة إلى الغد
٨	إفادتى من المحاضرات محدودة بسبب سوء التدريس
٩	يضايقنى ضعف ارتباط محتوى معظم المناهج الدراسية بحياتى العملية
١٠	معظم المحاضرات ثقيلة على نفسى بسبب طابعها التلقينى
١١	تصيبنى المحاضرات المتصلة أو المتتالية بدون راحة بالضجر
١٢	يؤرقنى عدم اهتمام المعلم بفهم الطالب لشرحه
١٣	يضايقنى كثرة التكليفات (بحوث ، تقارير ، امتحانات عملية ، ... إلخ)
١٤	يضايقنى صعوبة الكتب الدراسية المقررة
١٥	يضايقنى سوء طباعة الكتب الدراسية المقررة
١٦	يضايقنى تشدد بعض أعضاء هيئة التدريس فى منح التقديرات
١٧	يضايقنى اتباع بعض أعضاء هيئة التدريس أسلوباً مملاً فى التدريس
١٨	يضايقنى اعتماد الدراسة فى معظم المقررات الدراسية على الحفظ
١٩	يؤرقنى وجود مقررات دراسية ليس لها فائدة تطبيقية
٢٠	يضايقنى الشرح النظرى لمقررات عملية
٢١	يضايقنى خروج بعض أعضاء هيئة التدريس عن موضوع المحاضرة
٢٢	يؤرقنى تباعد أماكن المحاضرات المتتالية
٢٣	الكثير من استراتيجيات التدريس المستخدمة تسبب لى القهر والمعاناة
٢٤	استراتيجيات التدريس الخالية من المتعة والتشويق والإثارة تسهم فى تكوين

	اتجاه سالب نحو التعلم
٢٥	**محتوى معظم المقررات الدراسية جامد يخلو من الإمتاع**
٢٦	**أثناء الدراسة تنجرف أفكارى إلى مكان آخر لأن الأمر غاية فى الملل**
٢٧	**المقررات الدراسية فى تخصصى غاية فى الملل تجعلى منهك حتى فى التفكير فى الأمر**
٢٨	**أشعر بالتعب عند جلوسى على مقعدى لأنى أحس بالملل**
٢٩	**اعتماد التقويم على الامتحانات فقط يسبب لى ضيقاً شديداً**
٣٠	**قلة موضوعية أعضاء هيئة التدريس فى تقويم أداء الطلاب تثير مخاوفى**
٣١	**يضايقى كون الأساتذة فى واد وأنا فى واد آخر**
٣٢	**يضايقنى قلة التفاعل بين الطلاب وأعضاء هيئة التدريس**
م	**البنود**
٣٣	يضايقنى تباعد التعليم العالى عن سوق العمل
٣٤	يضايقنى قضاء وقتاً طويلاً متحيراً فى الإجابة عن الأسئلة
٣٥	ما أدرسه لا قيمة له فى حياتى المستقبلية
٣٦	الدراسة عملية مملة لأقصى حد
٣٧	ليس لدى رغبة فى التعلم لأنى أحس بالملل
٣٨	لا تستحق الدراسة المجهود الذى أبذله من أجلها
٣٩	أشعر بالملل عندما أدرس مقررات معينة
٤٠	**محتوى المقررات الدراسية غاية فى الملل حيث أجد نفسى فى أحلام اليقظة فى الكثير من الأحيان**
٤١	**غالباً أكون غير مدفوع لاستثمار جهدى فى المقررات الدراسية المملة**

ملحق (٣)

بنود استبيان الأسلوب المعرفى اللفظى / البصرى

م	**البنود**

١	أستمتع بأداء العمل الذى يتطلب استخدام الكلمات
٢	**أحلام يقظتى أحياناً تكون غاية فى التشويق لدرجة شعورى بأنى أعيشها فى الحقيقية**
٣	أستمتع بتعلم الكلمات الجديدة
٤	أستطيع بسهولة أن أفكر فى تحليل مرادفات الكلمات
٥	قوتى التخيلية أعلى من المتوسط
٦	نادراً ما أحلم
٧	أقرأ ببطء
٨	**لا أستطيع توليد صورة عقلية لوجه صديق عندما أغمض عيونى**
٩	**لا أعتقد أن بإمكان أى فرد أن يفكر باستخدام الصور العقلية**
١٠	**أفضل أن أقرأ التعليمات الخاصة بتنفيذ شىء ما فضلاً عن أن يشرحها لىّ شخص ما**
١١	أحلامى مشوقة بدرجة عالية جداً
١٢	طلاقتى فى استخدام الكلمات أعلى من المتوسط
١٣	أحلام يقظتى غير واضحة وغامضة
١٤	أقضى وقتاً قصيراً جداً فى محاولة زيادة مفرداتى اللغوية
١٥	يحتوى تفكيرى غالباً على الصور العقلية أو التخيلات
١٦	أكره البحث عن معانى الكلمات فى القواميس
١٧	أجد صعوبة كبيرة فى تذكر كلمات الأغانى
١٨	استخدم الرسوم التخطيطية فى شرح الأشياء
١٩	**أكره الرسومات والأشكال فى الكتب الدراسية أو المجلات أو الجرايد**
٢٠	استمتع بحل الكلمات المتقاطعة والسودوكو
٢١	أكره حل الألغاز المصورة
٢٢	**تساعدنى الرسوم التوضيحية أو الأشكال التخطيطية فى فهم ما أقرأه**

٢٣	أحفظ المعانى والمفاهيم الجديدة بسهولة
٢٤	أمتلك ذاكرة فوتوغرافية (صورية)
٢٥	أعتقد أن الصورة بألف كلمة
٢٦	أستطيع تصور ما يصفه الآخرون