التحقق من صلاحية أنموذج مقترح لتقييم جودة تقويم تعلمات الطلبة الجامعيين

Validation of a proposed model for evaluating the quality of university students' learning assessments

فاروق طباع*
أستاذ، جامعة سطيف2
Farouq Tebaa
Professor, University of Sétif2,
f.tebaa@univ-setif2.dz

يوسف خنيش
أستاذ، جامعة سطيف2
Youcef Khenniche
Professor, University of Sétif2
y.khenniche@univ-setif2.dz

تاريخ الاستلام: 2026/03/14 تاريخ القبول: 2026/05/11 تاريخ النشر: 2026/06/02

- **الملخص:** هدفت الدراسة الحالية إلى التحقق من صلاحية أنموذج مقترح لتقييم جودة تقويم تعلمات الطلبة الجامعيين. استنادا إلى الأدبيات ذات الصلة تم تحديد ستة معايير لتقييم جودة التقويم، تمثلت في الملاءمة، والأصالة، والشفافية، والإنصاف، وإمكانية التعميم، والعواقب التعليمية. وقد تم أجرأة كل معيار بمجموعة من المؤشرات التي تم تحويلها إلى عبارات لقياس منظورات الطلبة، ووُضعت في استبيان يتكون من (33) عبارة تم تطبيقه على عينة تكونت من (300) طالب وطالبة من جامعة سطيف2. تم فحص صلاحية الأنموذج باستخدام التحليل العاملي التوكيدي بتقدير النموذج بواسطة طريقة الأرجحية العظمى مع إعادة المعاينة (Bootstrap). توصلت النتائج إلى صلاحية الأنموذج المقترح لتقييم جودة التقويم استنادا إلى مؤشرات جودة المطابقة (X^2/df= 1,44; RMSEA= 0,038; RMR= 0,078; GFI= 0,885; AGFI= 0,865; CFI= 0,848; TLI= 0,833; IFI= 0,853)، حيث أظهرت مطابقة مقبولة للأنموذج المقترح مع بيانات العينة، مما يعزز تماسكه. كما أظهرت مستويات مقبولة من تشبعات العبارات على العوامل (0,36-0,59)، وترابطات العوامل (0,52-0,77) مما يدعم الصدق التقاربي والتمييزي للأنموذج. واستنادا إلى قيم "ألفا" و"أوميغا" و"الثبات المركب" (0,59-0,73) أظهرت تناسق عوامل الأنموذج. وبناء على النتائج المتوصل إليها تم مناقشة دلالاتها وفقاً للأدبيات النظرية والبحثية.

- **الكلمات المفتاحية:** أنموذج الجودة؛ معايير الجودة؛ التقويم الجامعي؛ منظورات الطلبة؛ طلبة الجامعة.

Abstract: The current study aimed to validate a proposed model for evaluating the quality of university students' learning assessments. Based on the relevant literature, six criteria were identified for assessing assessment quality, namely appropriateness, authenticity, transparency, fairness, generalizability, and educational consequences. Each criterion was implemented with a se t of indicators that were converted into items to measure students' perspectives. The items were included in a questionnaire consisting of 33 items, which was administered to a sample of 300 students from the University of Setif2. The validitation of the model was examined using confirmatory factor analysis (CFA) with the Maximum Likelihood (ML) method with bootstrapping. Based on the goodness-of-fit indices, the results confirmed the validity of the proposed model for assessing the quality of assessment (X^2/df= 1.44; RMSEA= 0.038; RMR= 0.078; GFI= 0.885; AGFI= 0.865; CFI= 0.848; TLI= 0.833; IFI= 0.853), showing acceptable fit of the

*- المؤلف المرسل

proposed model with the sample data, which reinforces its consistency. It also showed acceptable levels of factor loadings (0.36-0.59) and factor correlations (0.52-0.77), supporting the convergent and discriminant validity of the model. Based on the values of "Alpha," "Omega," and "Composite reliability" (0.59-0.73), the model factors were found to be consistent. Based on the findings, their implications were discussed in light of the theoretical and research literature.

Keywords: Quality model; quality criteria; university assessment; student perspectives; university students.

1- مقدمة:

يجتاز الطالب خلال مساره الجامعي العديد من التقويمات (اختبارات، استجوابات، مشاريع، عروض تقديمية، وغيرها) لجمع معلومات للاستخدام لتحسين الأداء والتقدم نحو الأهداف، وتدعيم النمو والتحسين المستمر، وتوثيق الأداء وإتقان المحكات لاتخاذ قرارات منح العلامات والانتقال والمصادقة على الكفاءة والمساءلة التعليمية (Stiggins, Arter, Chappuis & Chappuis, 2014). وحتى تكون تلك المعلومات ملائمة ودقيقة يُفترض أن تتمتع تلك التقويمات بالجودة في مختلف جوانبها؛ أسئلة الاختبار، والمهمات، وبرامج التقييم، وتنظيم التقويم وسياسته (Gerritsen-van Leeuwenkamp, Joosten-ten Brinke & Kester, 2018).

ومع تطوّر أساليب التقويم خلال العقود القليلة الماضية صحبتها تطورات في معايير فحص جودتها، لأن الاستناد إلى المعايير الكلاسيكية (الصدق والثبات) في فحص جودة تلك التقويمات غير كافية لأنها طوّرت في الأصل للاختبارات التقليدية الموضوعية والمقننة (Gipps, 1994)، لذلك اتجهت الجهود البحثية في الفترة الحالية نحو تطوير معايير جديدة لفحص جودة التقويمات تستند إلى الايدومترية الحديثة (Dochy, 2009).

وتمخضت عن تلك الجهود اقتراح معايير جديدة لجودة التقويم تتلاءم مع متطلبات التقويمات الحديثة، ومن بين تلك المعايير؛ الأصالة، والشفافية، والإنصاف، وإمكانية التعميم، والعواقب التعليمية، وهذا ما يسمح للقائمين على شؤون التعليم الجامعي بتوفير تقييمات عالية الجودة.

وتبعاً للأدبيات ذات الصلة تبيّن أن هناك أطر فكرية لمعايير جودة التقويم، من بينها الإطار الفكري الذي قدمه (Linn, Bakker & Dunbar, 1991) يشتمل على تسع معايير؛ العواقب، والإنصاف، والانتقال وإمكانية التعميم، والتعقيد المعرفي، وذات المعنى، وجودة المحتوى، وتغطية المحتوى، والتكاليف والفعالية. وقدم أيضا McMillan (1999; 2018) إطارا فكريا يستند إلى سبع معايير؛ وضوح وملاءمة الأهداف التعليمية، وملاءمة طرق التقويم، والصدق، والثبات، والعدالة، والنتائج الإيجابية، والفائدة والفعالية.

والإطار الفكري الذي لقي رواجا خلال العقد الماضي، ذلك الذي اقترحته (Baartman, Bastiaens, Kirschner, & Van der Vleuten, 2006; 2007)، يشتمل على 10 معايير وتم توسيعه إلى 12 معيارا؛ المقبولية، والأصالة، والتعقيد المعرفي، وقابلية المقارنة، والتكاليف والفعالية، والعواقب التعليمية، والإنصاف، والملاءمة للغرض، والملاءمة للتقييم الذاتي، وذات المعنى، وإعادة إنتاج القرارات، والشفافية، والعواقب التعليمية.

وبناء على الأطر الفكرية المتوافرة في الأدبيات، تبيّن وجود معايير عديدة لتقييم جودة التقويم (Linn et al., 1991; Baartman et al., 2006; 2007; McMillan, 1999; 2018; Tillema, Leenknecht & Segers, 2011; Gerritsen-van Leeuwenkamp et al., 2017; 2018)، وبعد مراجعتها وتحليلها تم استخلاص أنموذج لجودة التقويم الجامعي يتألف من ستة معايير تمثلت في الملاءمة، والأصالة، والشفافية، والإنصاف، وإمكانية التعميم، والعواقب التعليمية.

تشير **الملاءمة** إلى توافق أهداف البرنامج مع أنشطة وأساليب التقويم، وتغطية المعايير والمستويات للكفاءات، وملاءمة الأساليب المختلفة للتقويم (Baartman et al., 2007). فأغراض واستخدامات التقويم يجب أن ترتبط ارتباطا وثيقا بما يسعى إلى تحقيه من مستويات أو نواتج، لأن ذلك يساعد في الحصول معلومات توجه الأنشطة والقرارات المتعلقة بعملية التعلم (علام، 2004). وتعني **الأصالة** واقعية أنشطة ومهمات التقويم بحيث تسمح باستثارة الأداء في مواقف طبيعية، وجذابة ومفيدة ومثيرة للاهتمام بالنسبة للطالب، وتستثير المهارات المعرفية العليا، وتغطي مجالات محتوى التعلم الذي تلقاه الطالب (Tillema et al., 2011)

وترتبط **الشفافية** بمدى وضوح وقابلية فهم الطلبة وتنفيذهم لمهمات التقويم بالتعرف على الغرض من عملية التقويم ومعايير التصحيح والمكلفين بعملية التصحيح (Baartman, 2008)، بما يضمن تأثيرها على تعلم الطلبة وتحسين أدائهم بمعرفة معايير التصحيح (Dochy, 2009). أما **الإنصاف** فيتعلق بحصول جميع الطلبة على فرص متساوية للبرهنة على كفاءاتهم وتعزيز إمكانياتهم أثناء عملية التقويم (Dierick & Dochy, 2001). بما يضمن عدم وجود تحيّز لمجموعات معينة من الطلبة واستبعاد التباين غير المرتبط بمهمة التقويم (Tillema et al., 2011).

في حين أن تشير **إمكانية التعميم** إلى قابلية تعميم نتائج التقويم على نطاق أوسع من الطلبة والمهمات والمقوّمين وأساليب التقويم، بحيث لا يتوقف التقويم على أداء مهمة واحدة تُقوّم بمصحّح واحد خلال فترة واحدة، ولكن يجب أن تُعمّم على نطاق أوسع من المهمات والمصححين والفترات وأساليب التقويم (Scallon, 2004). ويضمن ذلك توفير ظروف ومعايير مماثلة لكل طالب، واتساق مستويات ومعايير التصحيح عبر الزمن وعبر المقوّمين، وإمكانية انتقال الأداء إلى مواقف أو مهمات أخرى(Tillema et al., 2011) . وتشير العواقب التعليمية إلى الأثار سواء

السلبية أو الإيجابية المترتبة من استخدام علمية التقويم (Messick, 1995). بمعنى أنها تشير إلى مدى تحقيق التقويم تأثيرات إيجابية وتقليل التأثيرات السلبية على التعلم والتدريس (Baartman et al., 2006). ويمكن أن تتضمن زيادة دافعية تعلم المفاهيم المهمة، وحل المشكلات، وتوظيف ما تعلمه الطالب، واستفادة المسؤولين من المعلومات لاتخاذ قرارات صائبة (Johnson, Penny & Gordon, 2009).

ويتم تقييم جودة التقويم الجامعي من منظور المستفيدين (أساتذة، طلبة، مديرين، مسيرين، أرباب العمل) التي تختلف باختلاف أدوارهم وخبراتهم (Baartman, Prins et al., 2011; Bronkhorst, Baartman, & Stokking, 2012) لأنها تزود بمعلومات مفيدة يمكن أن تسهم في تحسين الممارسات التقويمية. ويعدّ الطلبة من المستفيدين الذين يمكن أن يقدموا معلومات مفيدة حول جودة التقويم لأنهم المتأثرين مباشرة بجودة التقويم، فتصوراتهم وتوقعاتهم لجودة التقويم مهمة (Gerritsen-van Leeuwenkamp et al, 2018). فقد أظهرت دراسة Gerritsen-van Leeuwenkamp et al (2017) بأن مراعاة منظورات الطلبة في تقويم جودة التقويم ضرورية، وتعتبر المعلومات التي يقدمونها للأساتذة والمستشارين التعليميين والمديرين مهمة لاستخدامها في مراقبة جودة التقويم في الجامعة.

قام بعض الباحثين بفحص صلاحية الأطر الفكرية المقدمة وتطبيقها في جمع معلومات عن جودة التقويم في التعليم الجامعي (طباع وفخرو ومسعود، 2024)، والتعليم المهني العالي (Baartman et al., 2006; 2007)، وتعليم الصيدلة (Shah, McLaughlin, Eckel, Mangun & Hawes, 2016)، والتعليم الرياضي (Borghouts, Slingerland & Haerens, 2017)، والتعليم والتدريب المهني (Gulikers, Biemans & Mudler, 2009)، وفي المدارس (Plake, Impara & Bukendahl, 2004).

واستُخدمت في فحص من صلاحية الأطر الفكرية طرق متنوعة، فقد اعتمدت (Baartman et al 2006) على خبراء التقويم باستخدام طريقة كيفية قائمة على المجموعة المركزية، وفي الفترة نفسها تقريبا أجرى (Baartman et al., 2007) مقارنة كيفية لإطارين فكريين لتقييم جودة تقويم الكفاءة، تضمن الأول (10) معايير طوّرت من قبلهم والإطار الفكري لصدق المفهوم لـ "ميسيك" Messick، مما سمح بإضافة معايير إضافية جديدة. وأجرى (Tillema et al., 2011) مراجعة كيفية للدراسات لتحديد معايير جودة تقويم الأقران بما يتلاءم مع مراحل دورة التقويم، وأفرزت عن مطابقة معايير الجودة لمراحل دورة التقويم. واستندت مراجعة (Gerritsen-van Leeuwenkamp et al., 2017) إلى التحليل المفاهيمي للنص لتقديم إطار لمعايير الجودة،

واستندت (Gerritsen-van Leeuwenkamp et al., 2018) إلى التحليل العاملي الاستكشافي لتحديد مكونات جودة التقويم لبناء استبيانات لقياس تصورات وتوقعات الطلبة لجودة التقويم.

وعلى الرغم من توافر نماذج لتقييم جودة التقويم الجامعي من منظور الطلبة، إلا أنه ما تزال محدودة فيما يتعلق بالتحقق من صلاحية النماذج المقترحة، فقد اكتفت بالأساليب الكلاسيكية القائمة على التحليل الكيفي أو التحليل العاملي الاستكشافي، دون اختبار مدى ملاءمتها عبر مؤشرات جودة المطابقة التي تستند إلى التحليل العاملي التوكيدي.

التحليل العاملي التوكيدي طريقة إحصائية متقدمة تستخدم للتأكد من ملاءمة نموذج القياس المستخدم مع بيانات العينة (Byrne, 2016). وتعتمد الطريقة على الحكم على مطابقة النموذج العاملي الذي يعكس بنية المفهوم وفقاً للعديد من الأدلة، ومقارنة النماذج المتنافسة بمرونة بناء على مؤشرات جودة المطابقة (تيغزة، 2017).

وبالتحديد فان الفجوة البحثية هنا تكمن في الحاجة إلى اختبار أنموذج تقييم متكامل لجودة التقويم الجامعي من منظور الطلبة، والتحقق من خصائصه القياسية عبر إجراءات تحليلية متقدمة تشمل اختبار الصدق البنائي عن طريق التحليل العاملي التوكيدي بما يوفره كذلك من مؤشرات للصدق التقاربي والتمييزي، واختبار الاتساق باستخدام مؤشرات الثبات. هذا يسهم في توفير أداة قياس يمكن الاعتماد عليها في تقييم وتحسين أنظمة التقويم الجامعي واتخاذ قرارات تطويرية تستند إلى بيانات موثوقة.

وعليه، تسعى هذه الدراسة إلى التحقق من صلاحية الأنموذج المقترح لتقييم جودة التقويم الجامعي من منظور الطلبة، من خلال اختبار بنيته العاملية وملاءمته لبيانات العينة، وتقييم مؤشرات الصدق والثبات المرتبطة به، بما يدعم موثوقية نتائجه وقابليته للتطبيق في المؤسسات الجامعية. وذلك بالإجابة على التساؤل التالي: **ما مدى صلاحية (صدق وثبات) الأنموذج المقترح لتقييم جودة التقويم الجامعي من منظور الطلبة؟**

2- الطريقة والأدوات:

2-1- العينة: أجريت الدراسة الحالية بجامعة سطيف2، موزعة على ثلاث كليات (العلوم الإنسانية والاجتماعية، وكلية الحقوق والعلوم السياسية، وكلية الآداب واللغات)، والتي توزعت بدورها على الأقسام والمستويات الدراسية. حيث تكونت العينة من 300 طالب وطالبة، تتراوح أعمارهم بين 18 و43 سنة، بمتوسط (22,34) سنة، وانحراف معياري (2,88). حيث يتوزع أفراد العينة حسب الجنس، والمستوى، والكلية، والقسم، كما هو موضح في الجدول1.

الجدول 1. توزيع أفراد العينة حسب متغيرات الجنس والمستوى والكلية والقسم

المتغيرات	الفئات	عدد الأفراد	النسبة %
الجنس	ذكر	97	32,3
	أنثى	203	67,7
	المجموع	300	100,0
المستوى الدراسي	السنة الأولى	46	15,3
	السنة الثانية	62	20,7
	السنة الثالثة	64	21,3
	ماستر1	43	14,3
	ماستر2	85	28,3
	المجموع	300	100,0
الكلية	كلية العلوم الإنسانية والاجتماعية	175	58,3
	كلية الآداب واللغات	74	24,7
	كلية الحقوق والعلوم السياسية	51	17,0
	المجموع	300	100,0
القسم	علم النفس وعلوم التربية والأرطفونيا	64	21,3
	علم الاجتماع	47	15,7
	التاريخ والأثار	17	5,7
	الإعلام والاتصال	14	4,7
	الفلسفة	14	4,7
	اللغة الإنجليزية	24	8,0
	اللغة الفرنسية	23	7,7
	اللغة والأدب العربي	27	9,0
	الحقوق	33	11,0
	العلوم السياسية	18	6,0
	علوم اجتماعية	19	6,3
	المجموع	300	100,0

يوضح الجدول1 أن نسبة الذكور (32.3%) أقل بكثير من الإناث (67.7%). أما حسب المستوى الدراسي من الأولى ليسانس إلى الثانية ماستر تراوحت نسبتهم بين (14.3-28.3%)، حيث سُجلت أعلى نسبة في الثانية ماستر (28,3%)، وأدنى نسبة في الأولى ماستر (14,3%)). أما حسب الكليات فجاءت أعلى نسبة بكلية العلوم الإنسانية والاجتماعية بـ (58.3%) ، تليها كلية الآداب

واللغات بـ (24.7%)، ثم كلية الحقوق والعلوم السياسية بـ (17%). وفيما يتعلق بالأقسام فتراوحت النسب بين (4.7-21.3%)، بنسبة أعلى في قسم علم النفس وعلوم التربية والأرطفونيا (21.3%)، وأدنى نسبة في قسم الإعلام والاتصال بـ (4.7%).

2-2- أداة جمع البيانات:

من الناحية النظرية، يتضمن الأنموذج المقترح لتقييم جودة التقويم الجامعي ست معايير (الملاءمة، الأصالة، الشفافية، الإنصاف، إمكانية التعميم، العواقب التعليمية) ، تم استخلاصها من بعض أدبيات جودة التقويم (Baartman & al, 2006; Gerritsen-van Leeuwenkamp & al, 2017; 2018; McMillan, 1999; 2018; Tillema & al, 2011) يشتمل كل معيار على عينة من المؤشرات التي تم صياغتها في شكل عبارات تم إدراجها في الاستبيان. حيث أن المفهوم المقاس في الأنموذج يتعلق بجودة التقويم الجامعي، ويشتمل المفهوم المقاس ستة عوامل (متغيرات كامنة)، وكل عامل تم أجرأته في عينة من العبارات. حيث يوضح الشكل1 النموذج النظري المقترح لتقييم جودة التقويم الجامعي من منظور الطلبة، والذي تم على أساسه إعداد الاستبيان.

ومن الناحية الفنية، تم إعداد استبيان لقياس تصورات الطلبة لجودة التقويم الجامعي بما يتوافق مع الإطار الفكري المقترح. اشتمل الاستبيان على (33) عبارة موزعة على ستة معايير، كما يلي؛ الملاءمة (5 عبارات: من 1 إلى 5)، والأصالة (5 عبارات: من 6 إلى 10)، والشفافية (5 عبارات: من 11 إلى 15)، والإنصاف (5 عبارات: من 16 إلى 20)، وإمكانية التعميم (5 عبارات: من 21 إلى 25)، والعواقب التعليمية (8 عبارات: من 26 إلى 33). يتم الإجابة على العبارات وفق سلم ليكرت الخماسي ("تنطبق تمامًا" إلى "لا تنطبق إطلاقًا")، وتُعطى لكل إجابة من الإجابات درجات تتراوح بين (5) و(1) درجات.

3-2- الحدود الزمانية والمكانية:

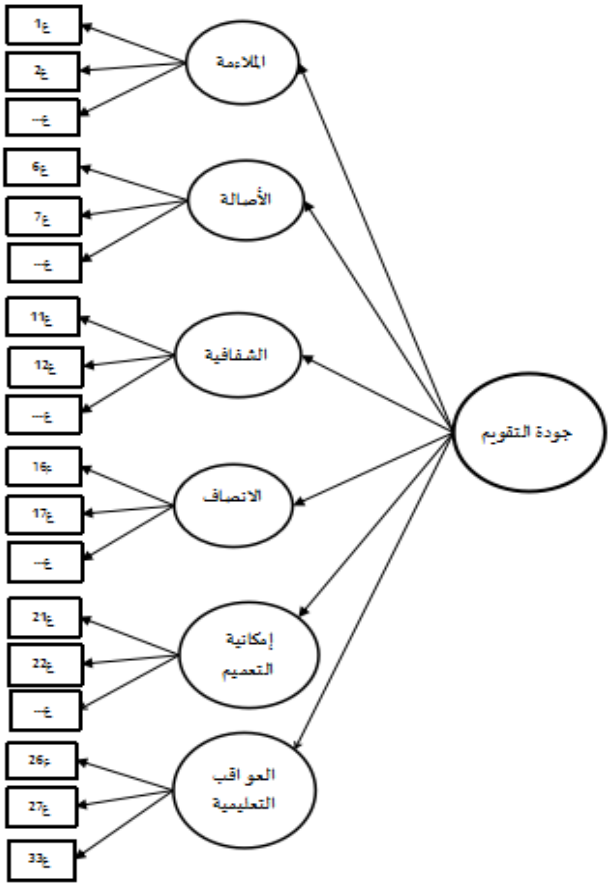
تم تطبيق الدراسة بالكليات الثلاث التي تتضمنها جامعة سطيف2 (كلية العلوم الإنسانية والاجتماعية، كلية الحقوق والعلوم السياسية، كلية الآداب واللغات)، وفي كل الأقسام التي تتضمنها الكليات. وأجريت الدراسة في الفترة الممتدة بين 30 أفريل 2024 و14 ماي 2024.

4-2- الأساليب الإحصائية:

تم استعمال برمجية SPSS AMOS. 24 لإجراء التحليل العاملي التوكيدي للتحقق من صلاحية الأنموذج المفترض، حيث تم حساب قيم الالتواء والتفلطح والنسب الحرجة لها لاستكشاف البيانات، وتقدير مؤشرات جودة المطابقة بطريقة الأرجحية العظمى (مربع كاي χ^2 ، نسبة مربع كاي χ^2/df، الجذر التربيعي لمتوسط الخطأ التقريبي RMSEA، جذر متوسط مربعات

البواقي RMR، مؤشر جودة المطابقة GFI، مؤشر جودة المطابقة المصحح AGFI، مؤشر المطابقة المقارن CFI، مؤشر توكر- لويس TLI، مؤشر المطابقة التزايدية IFI) لتقييم مدى ملائمة النموذج المفترض مع بيانات العينة، وقد تم توظيف طريقة إعادة المعاينة Bootstrap لتقدير النموذج والاختبارات الإحصائية لضمان الثقة والدقة في النتائج.

كما تم تقدير مصفوفة التشبعات المعيارية للعبارات لتحليل العلاقات بين العبارات والعوامل الستة، ومصفوفة الارتباطات بين العوامل للتأكد من الارتباطات بين العوامل المكونة للأنموذج. بالإضافة إلى تقدير الثبات باستخدام معاملات "ألفا" و "أوميغا" و"الثبات المركب" لفحص مدى اتساق وموثوقية درجات العوامل.



الشكل1. الأنموذج النظري المقترح لتقييم جودة التقويم الجامعي

3- النتائج:

قبل التحقق من جودة مطابقة النموذج المفترض مع بيانات العينة، تم فحص اعتدالية توزيع متغيرات النموذج من خلال قيم الالتواء والتفلطح والنسبة الحرجة لكل منهما، وذلك للتعرف على مدى توفر افتراضات طريقة الأرجحية العظمى.

الجدول 2. مؤشرات الاعتدالية (الالتواء والتفلطح) لمتغيرات الأنموذج

رقم العبارة	معامل الالتواء	النسبة الحرجة c. r.	معامل التفلطح	النسبة الحرجة c. r.
1	-1,136	-8,033*	1,006	3,556*
2	-,619	-4,379*	-,218	-,769
3	-,809	-5,718*	-,085	-,300
4	-,537	-3,799*	-,633	-2,238*
5	-,406	-2,868*	-,701	-2,479*
6	-,221	-1,566	-,987	-3,490*
7	-,398	-2,815*	-,837	-2,958*
8	-,312	-2,207*	-,937	-3,313*
9	-,304	-2,150*	-1,112	-3,932*
10	-1,080	-7,639*	,414	1,463
11	-,387	-2,733*	-,887	-3,134*
12	-,363	-2,566*	-,774	-2,737*
13	-,305	-2,154*	-1,007	-3,559*
14	-,343	-2,428*	-,719	-2,542*
15	-,494	-3,491*	-,725	-2,562*
16	-,471	-3,331*	-,580	-2,052*
17	-,370	-2,614*	-,833	-2,945*
18	-,424	-3,000*	-,627	-2,217*
19	-,597	-4,219*	-,659	-2,328*
20	-,452	-3,199*	-,964	-3,409*
21	-,606	-4,283*	-,629	-2,223*
22	-,337	-2,384*	-,946	-3,346*
23	-,447	-3,164*	-,749	-2,649*
24	-1,288	-9,107*	1,003	3,548*

25	-,639	-4,517*	-,379	-1,339
26	-0,597	-4,219*	-0,659	-2,328*
27	-0,606	-4,283*	-0,629	-2,223*
28	-,490	-3,466*	-,949	-3,356*
29	-,532	-3,760*	-,750	-2,652*
30	-,552	-3,901*	-,549	-1,940
31	-,576	-4,075*	-,625	-2,210*
32	-,399	-2,818*	-,812	-2,870*
33	-,580	-4,104*	-,555	-1,962*
التفلطح متعدد المتغيرات * دالة عند 0,05			119,316	21,499

تُظهر نتائج اختبار الاعتدالية (الجدول 2) أن معظم قيم الالتواء والتفلطح للمتغيرات دالة إحصائيا، حيث أن غالبية القيم أكبر من النسبة الحرجة (c.r > ±1.96)، وهذا يدل على انتهاك افتراض الاعتدالية الأحادية للمتغيرات. كما تُظهر قيمة التفلطح متعددة المتغيرات (c.r. = 21,499) انتهاكا للاعتدالية متعددة المتغيرات.

بناء على نتائج اختبار الاعتدالية يتبين بأن معظم العبارات تنحرف عن التوزيع الاعتدالي، هذا الانتهاك لافتراضات طريقة التقدير التقليدية للأرجحية العظمى Maximum Likelihood يبرر استخدام إعادة المعاينة Bootstrap في تقدير النموذج والاختبارات الإحصائية، لضمان دقة النتائج والحصول على فترات ثقة ذات موثوقية وتجنب التحيز في النتائج (Byrne, 2016).

تم تقدير مؤشرات جودة المطابقة الأكثر شيوعاً واستخداما في الأدبيات البحثية استناداً إلى توصيات وإرشادات (Byrne, 2016)، حيث أنه تم مؤشرات جودة المطابقة؛ مربع كاي χ^2، نسبة مربع كاي χ^2 /df، الجذر التربيعي لمتوسط الخطأ التقريبي RMSEA، جذر متوسط مربعات البواقي RMR، مؤشر جودة المطابقة GFI، مؤشر جودة المطابقة المصحح AGFI، مؤشر المطابقة المقارن CFI، مؤشر توكر- لويس TLI، مؤشر المطابقة التزايدية IFI.

أظهرت نتائج التحليل (جدول3) أن قيمة مربع كاي كانت دالة (χ^2 = 691.260, df = 480, p < .001)، وهي متوقعة نظرًا لحساسية المؤشر لحجم العينة. إلا أن نسبة مربع كاي أقل من 3 (χ^2/df =1.44)، وهذا يشير إلى مطابقة جيدة للنموذج. كما بلغت قيمة الجذر التربيعي لمتوسط الخطأ التقريبي (0.038) RMSEA وهي ضمن المجال المثالي، وكانت قيمة جذر متوسط مربعات البواقي (0.078) RMRضمن الحدود المقبولة. بينما جاءت مؤشرات المطابقة المقارنة؛ المطابقة

التزايدية (IFI= 0.853)، وتوكر- لويس (TLI= 0.833)، والمطابقة المقارن (CFI= 0.848) دون الحد المقبول تقليديًا (0.90) إلاّ أنها قريبة، كما هو الحال بالنسبة لقيمة مؤشر جودة المطابقة (GFI= 0.885)وجودة المطابقة المصحح (AGFI= 0.865) التي جاءت أقل بقليل من المستوى المعياري المقبول.

الجدول 3. مؤشرات جودة المطابقة للأنموذج المفترض لجودة التقويم الجامعي

المؤشرات	القيمة	شروط القبول
مؤشر مربع كاي χ^2	χ^2= 691,260; df= 480 (p= 0,000)	تكون غير دالة
مؤشر نسبة مربع كاي χ^2 /df	1,44	3 >
مؤشر الجذر التربيعي لمتوسط الخطأ التقريبي RMSEA	0,038	0,08 ≥
جذر متوسط مربعات البواقي RMR	0,078	0,08 ≥
مؤشر جودة المطابقة GFI	0,885	0,90 ≤
مؤشر جودة المطابقة المصححAGFI	0,865	0,90 ≤
مؤشر المطابقة المقارن CFI	0,848	0,90 ≤
مؤشر توكر - لويس TLI	0,833	0,90 ≤
مؤشر المطابقة التزايدية IFI	0,853	0,90 ≤

بوجه عام، تؤكد النتائج بأن النموذج المفترض يتمتع بجودة مطابقة من جيدة إلى مقبولة مع بيانات العينة. كما أن نتائج Bootstrap تؤكد استقرار النموذج وخلوه من المشاكل. واستناداً إلى تقييم المؤشرات مجتمعة فان النموذج يتمتع بمطابقة مقبولة تسمح بتفسير العلاقات بين المتغيرات، مما يشير إلى صلاحية النموذج النظري المقترح لتقييم جودة التقويم الجامعي.

الجدول 4. قيم التشبعات المعيارية للعبارات على العوامل

رقم العبارة	العبارة	قيمة التشبع
1	مواضيع وأنشطة التقييم والاختبارات تتلاءم مع أهداف البرنامج الدراسي	0,483
2	معايير تصحيح الاختبارات والتقييمات لها صلة بالأهداف الدراسية	0,530
3	أنشطة التقويم وأسئلة الاختبارات لها صلة بعناصر محتوى البرنامج الدراسي.	0,474
4	تقيس الاختبارات والتقييمات مختلف عملياتي المعرفية (الفهم، التحليل، التقويم...)	0,496
5	تُعطى لي إمكانية لصياغة أهداف تعلّمي بنفسي	0,425
6	الاختبارات والتقييمات التي أجتازها تشبه وضعيات الحياة المهنية المستقبلية	0,512
7	أسئلة الاختبارات والتقييمات ذات معنى وقيمة بالنسبة لي	0,412
8	معايير تصحيح الاختبارات والتقييمات مشابهة لمعايير المهنة المستقبلية	0,588

9	عمليات التفكير التي أقيم عليها في الاختبارات نفسها عمليات التفكير في المهنة	0,525
10	ظروف إنجازي للتقييمات والاختبارات تشبه ظروف العمل التي سأصادفها في المهنة	0,441
11	يتم إعلامي بالمطلوب مني إنجازه في الاختبارات حتى أكون قادرا على ضبط أدائي	0,473
12	تقدم لي تغذية راجعة مفصلة بشأن نتائجي في الاختبارات والتقييمات	0,454
13	يتم إعلامي مسبقا بأي تغييرات قد تطرأ على طريقة تقييمي	0,445
14	أسئلة الاختبارات والتقييمات قابلة للإنجاز (غير تعجيزية)	0,533
15	أكون على علم بالغرض من التقييمات (لتعزيز تعلمي أو لمنحي العلامة أو كلاهما)	0,439
16	يحصل جميع الطلبة على فرص متكافئة لإظهار كفاءتهم فيالاختبارات والتقييمات	0,444
17	أثناء الاختبارات والتقييمات لا توجد عوامل تؤثر على أداء الطلبة كالغش والضجيج	0,390
18	تعليمات الإجابة على الاختبارات والتقييمات واضحة لدى جميع الطلبة	0,406
19	المدة المحددة للإجابة على الاختبارات والتقييمات كافية لجميع الطلبة	0,387
20	أثناء تصحيح الاختبارات تستخدم نفس معايير التصحيح على كل الطلبة	0,362
21	العلامة التي تمنح لي في الاختبارات والتقييمات تعبر فعلا عن تمكني من المقياس	0,422
22	سوف أحصل على نفس العلامة إذا أعيد اختبار أو تقييم أخر من نفس الموضوع	0,428
23	سوف أحصل على نفس العلامة إذا أعيد الاختبار أو التقييم في فترة أخرى	0,476
24	يتم تقييمي باستخدام أسئلة ومهمات متنوعة لإعطائي فرصة لإظهار قدراتي	0,443
25	يتم منحي الفرصة للمناقشة مع أستاذي وزملائي في وضع علاماتي	0,485
26	تعزز التقييمات والاختبارات رغبتي أكثر في التعلم والتحسين	0,479
27	تساعد الاختبارات والتقييمات الأستاذ على تعديل وتحسين طرق ووسائل التدريس	0,405
28	تساعدني الاختبارات والتقييمات على استثمار وتوظيف ما تعلمته في الواقع	0,518
29	تساعدني نتائج التقويم على تحديد نقاط القوة والضعف في تعلماتي	0,534
30	تعطي لي الاختبارات والتقييمات الفرصة لاستدراك الضعف في نتائجي لتحسين أدائي	0,553
31	أسئلة الاختبارات والتقييمات تعطي لي فرصة لتعزيز وتحسين تعلمي	0,531
32	تساعدني الاختبارات والتقييمات في الاحتفاظ على الكفاءة المكتسبة في المستقبل	0,515
33	تشجعني الاختبارات والتقييمات على التعلم الذاتي والبحث المستقل عن المعارف	0,437

يوضح الجدول 4 أن معظم القيم المعيارية لتشبعات العبارات على عواملها تقع بين (0,40-0,60)، حيث تعدّ تشبعات مقبولة لأنها أكبر من (0,40)، باستثناء العبارات رقم (17، 19، 20) في عامل "الإنصاف" التي جاءت قريبة (0,40). وقد تراوحت قيم تشبعات العبارات على العوامل التي تندرج ضمنها بين (0,362-0,588). تشير قيم التشبعات العاملية إلى قوة العلاقة بين العبارات والعوامل الكامنة التي تنتمي إليها. وقد جاءت معظم التشبعات مرتفعة، وهذا يدل على أن العبارات تمثل أبعادها النظرية تمثيلًا جيدًا، ويدعم بذلك الصدق التقاربي للأنموذج.

وتم التحقق من الارتباطات البينية بين متغيرات الأنموذج المفترض لتقييم جودة التقويم من خلال التباينات المشتركة بين العوامل للتحقق من مدى ابتعاد الأنموذج المقترح عن النموذج الصفري، حيث لا يجب أن تتعدى الارتباطات بين العوامل المدى (0,20- 0,90).

الجدول 5. مصفوفة الارتباطات بين العوامل المكونة للأنموذج

العوامل	(1)	(2)	(3)	(4)	(5)	(6)
الملاءمة (1)	1,00	0,653	0,582	0,734	0,522	0,515
الأصالة (2)	---	1,00	0,624	0,735	0,683	0,547
الشفافية (3)	---	---	1,00	0,771	0,716	0,516
الإنصاف (4)	---	---	---	1,00	0,744	0,698
إمكانية التعميم (5)	---	---	---	---	1,00	0,591
العواقب التعليمية (6)	---	---	---	---	---	1,00

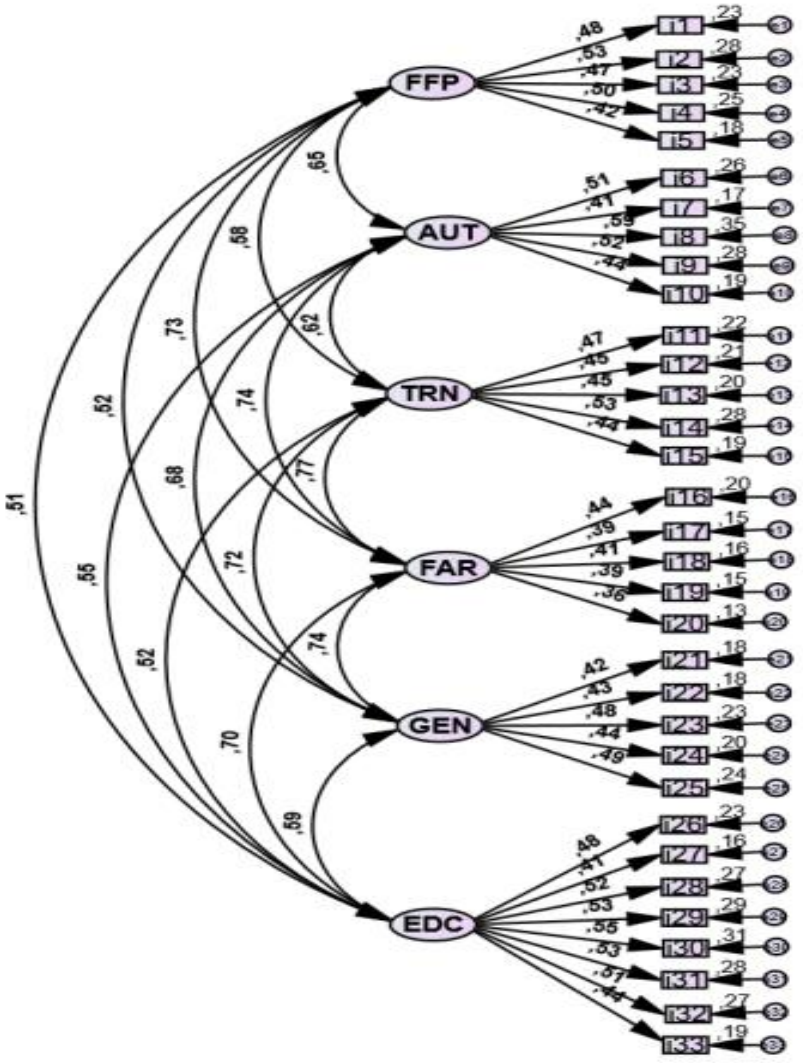
تشير نتائج معاملات الارتباط (الجدول5) إلى وجود علاقات موجبة من متوسطة إلى مرتفعة بين جميع العوامل الكامنة، وهذا يتوافق مع النموذج النظري الذي يفترض وجود ترابط بنيوي بين العوامل. وفي الوقت نفسه، فإن عدم تجاوز الارتباطات للحدود الحرجة (0.85) يدل على احتفاظ كل عامل بقدر من الاستقلال النسبي، وهذا يدعم البنية متعددة الأبعاد للأنموذج. وهذا مؤشر دال على الترابط المفاهيمي بين العوامل المكونة للأنموذج وتمتعه بالصدق التمييزي. وقد تم عرض نتائج الجدول 4 المتعلق بقيم تشبعات العبارات على عواملها، ونتائج الجدول5 المتعلق بالارتباطات البينية (التباينات المشتركة) بين متغيرات الأنموذج المقترح في الشكل 5.

الثبات: تم التحقق من ثبات درجات العوامل المكونة لأنموذج تقييم جودة التقويم الجامعي باستخدام معامل "ألفا كرونباخ" ومعامل "أوميغا ماكدونالد"، ومعامل "الثبات المركب" للتناسق الداخلي بين عبارات كل عامل.

الجدول6. معاملات الثبات "ألفا"، و"أوميغا" و"الثبات المركب" للعوامل

العوامل	عدد العبارات	المتوسط الحسابي	الانحراف المعياري	معامل ألفا α	معامل أوميغا ω	الثبات المركب CR
الملاءمة	5	3,76	0,68	0,60	0,61	0,60
الأصالة	5	3,49	0,79	0,62	0,63	0,62
الشفافية	5	3,40	0,77	0,56	0,57	0,56
الإنصاف	5	3,57	0,69	0,59	0,60	0,59
إمكانية التعميم	5	3,64	0,73	0,60	0,59	0,61
العواقب التعليمية	8	3,67	0,70	0,73	0,74	0,72

أظهرت نتائج تحليلات الثبات (جدول6) أن معاملات الثبات ألفا، وأوميغا، والثبات المركب متقاربة عبر جميع العوامل، وهذا يعد مؤشرا إيجابيا على استقرار تقديرات الثبات وعدم تأثرها بـ طريقة الحساب المعتمدة. مما يشير إلى تجانس العبارات داخل كل عامل. أعلى معامل للثبات (ألفا ، أوميغا، الثبات المركب) كان في عامل العواقب التعليمية (0,73 و0,74 و0,72) الذي يعدّ ضمن المـ ستوى الجيد.



FFP : الملاءمة، AUT : الأصالة، TRN : الشفافية، FAR : الإنصاف، GEN : إمكانية التعميم، EDC : العو اقب التعليمية

الشكل2. الرسم التخطيطي للنموذج العاملي المفترض متضمنا البارامترات المقدرة

أما باقي العوامل (الملاءمة، الأصالة، الشفافية، الإنصاف، إمكانية التعميم) فقد تراوحت معاملات الثبات فيها بين (0,56- 0,62)، والتي تعد أقل من الحد المثالي (0,70)، ولكنها تعد مقبولة في سياق التكوينات الفرضية متعددة العوامل أو في حالة قلة عدد العبارات.

4- المناقشة:

أظهرت نتائج التحليل العاملي التوكيدي باستخدام طريقة الأرجحية العظمى مع إعادة المعاينة أن الأنموذج المقترح يتمتع بدرجة مطابقة مقبولة مع بيانات العينة. تشير هذه النتائج إلى أن النموذج السداسي الأبعاد يفسر البنية الكامنة لجودة التقويم بدرجة مقبولة، وإن لم يصل إلى مستوى المطابقة المثالية، لأنه متعدد الأبعاد (ستة معايير مترابطة)، وحساسية مؤشرات المطابقة لحجم العينة وتعقيد النموذج، وطبيعة البيانات التصورية المعتمدة على تقديرات الطلبة، لذلك يمكن القول إن النتائج تدعم مصداقية وصلاحية الأنموذج.

تتسق نتائج الدراسة الحالية مع الاتجاهات الحديثة لأدبيات تقييم جودة التقويم (Baartman et al.,2006; Gerritsen-van Leeuwenkamp et al., 2018; Tillema et al., 2011) التي تنظر إليه كمفهوم متعدد الأبعاد، وليس كمفهوم أحادي يرتبط فقط بصدق الاختبار أو بثباته. حيث يُنظر إلى جودة التقويم بنظرة أكثر شمولية وأكثر اتساعا نظرا لتكامل المعايير الكلاسيكية والمعايير الحديثة، وتكامل أساليب التقويم التقليدية وأساليب التقويم الجديدة.

رغم أن تشبعات بعض العبارات جاءت ضمن الحد الأدنى المقبول (0,30- 0,40) إلا أن جميعها دالة تفوق الحد الأدنى المعتمد في الدراسات الاجتماعية، وهذا يدعم الصدق التقاربي لأن المؤشرات تعكس تمثيلا معقولا لبعدها الكامن، على الرغم أن الأنموذج قابلا للتحسين في بعض عباراته.

الارتباطات بين العوامل جاءت من متوسطة إلى مرتفعة باعتبار أنها لم تتجاوز (0,85) ولم تقل عن (0,20)، مما يدل على عدم وجود تداخل بين الأبعاد. وهذا يؤكد على تحقق الصدق التمييزي، ويعكس الطبيعة المتكاملة لمعايير جودة التقويم، فالملاءمة والأصالة والشفافية والإنصاف وإمكانية التعميم والعواقب التعليمية أبعاد مترابطة نظريا ولكنها غير متطابقة. وقد أكد McMillan (1999) أن جودة التقويم تتكون من أبعاد مترابطة تشمل العدالة والشفافية وملاءمة المهام، وهي أبعاد متقاربة مفاهيمياً مع معايير الأنموذج المقترح.

وتراوحت قيم الثبات بين (0,59- 0,73)، وتطابق مؤشرات (ألفا وأوميغا والثبات المركب) مما يعزز موثوقية النتائج، ويشير إلى استقرار البنية العاملية للأنموذج. وهذه النتائج تتوافق مع ما توصلت إليه دراسة (Baartman et al., 2007) التي كشفت عن معاملات ثبات مقبولة تراوحت بين (0,59 -0,82).

الأنموذج السداسي المقترح (الملاءمة، الأصالة، الشفافية، الإنصاف، إمكانية التعميم، العواقب التعليمية) يتسق مع أدبيات جودة التقويم الحديثة، خاصة التوجه الذي ينظر إلى جودة التقويم بوصفه مفهوما متعدد الأبعاد، الذي يتجاوز المفهوم التقليدي ثنائي البعد (الصدق والثبات). ويمكن تفسير ترابط الأبعاد إلى ترابط مراحل دورة التقويم (Tillema et al., 2011). حيث أن الملاءمة والأصالة تعكسان جودة تصميم المهمات والأنشطة وملاءمتها لأغراض التقويم، وترتبط الشفافية والإنصاف بإجراءات تطبيق المهمات، وتعكس إمكانية التعميم اتساق عملية التصحيح والتقدير، وتمثل العواقب التعليمية التأثيرات الناجمة عن استخدام التقويم كالتغذية الراجعة والدافعية وتحسين التعلم. وهذا التكامل يدعم البناء المفاهيمي للأنموذج ويعزز أهميته التطبيقية في التعليم الجامعي.

تشير النتائج إلى أن النموذج المقترح يتمتع بدرجة مقبولة من الصلاحية البنائية والاتساق الداخلي، وهذا يسمح باعتماده إطاراً مرجعياً لتقييم جودة التقويم الجامعي من منظور الطلبة. ورغم أن بعض المؤشرات لم تبلغ مستوى المطابقة المثالية، فإن تكامل الأدلة الإحصائية (المطابقة، التشبعات، الترابطات، الثبات) يدعم تماسك النموذج وقابليته للتطبيق، مع إمكانية تحسينه مستقبلاً عبر تنقيح بعض مؤشراته واختباره في سياقات أوسع.

5- خاتمة وتوصيات:

هدفت الدراسة إلى التحقق من صلاحية أنموذج مقترح لتقييم جودة تقويم الجامعي في ضوء ستة معايير؛ الملاءمة، الأصالة، الشفافية، الإنصاف، إمكانية التعميم، والعواقب التعليمية. وقد أظهرت نتائج التحليل العاملي التوكيدي باستخدام الأرجحية العظمى مع إعادة المعاينة أن الأنموذج يتمتع بدرجة مطابقة مقبولة مع بيانات العينة، مما يدعم صلاحيته البنائية. وأشارت مؤشرات المطابقة إلى أن البنية السداسية تعكس تصوراً متماسكاً لجودة التقويم من منظور الطلبة. ودعمت التشبعات العاملية المقبولة والارتباطات المتوسطة بين الأبعاد تحقق الصدق التقاربي والتمييزي، كما أظهرت معاملات الثبات مستويات مقبولة من الاتساق الداخلي.

وتتسق هذه النتائج مع التصور المعاصر لمفهوم الصلاحية كما بلوره Messick (1995) الذي يؤكد أن جودة التقويم تتجاوز الدقة القياسية التي تستند إلى الصدق والثبات لتشمل العواقب التعليمية والأخلاقية، كما تنسجم مع فلسفة التقويم الذي يُنظر إليه بوصفه أداة لتحسين التعلم وليس مجرد وسيلة لإصدار الحكم.

وعليه، يمكن القول إن هذه الدراسة ساهمت في بناء إطار مرجعي لتقييم جودة التقويم الجامعي من منظور الطلبة، يجمع بين البعد القياسي، والتربوي، والأخلاقي في نموذج بنائي متكامل. وبناء على النتائج التي توصلت إليها الدراسة يمكن تقديم المقترحات والتوصيات التالية:

- إعادة اختبار الأنموذج في سياقات أخرى وتنقيح العبارات ذات التشبعات المنخفضة لتحسين جودة المطابقة، ومقارنة الأنموذج المقترح بنماذج بديلة لاختبار قوته التفسيرية.
- اقتراح نماذج أخرى لتقييم جودة التقويم الجامعي والتحقق من صلاحيتها من منظورات الفاعلين الآخرين كالأساتذة، والإداريين.
- اعتماد الأنموذج المقترح كأداة لتقييم جودة التقويم في الكليات والجامعات لتحسين التقويمات الصفية.
- إشراك الطلبة والفاعلين الآخرين من أساتذة وإداريين في عمليات تحسين منظومة التقويم باعتبارهم شركاء في تقييم جودة التقويم الجامعي.

- قائمة المراجع:

تيغزة، أ. ب. (2011). *التحليل العامل الاستكشافي والتوكيدي*. عمان: دار المسيرة.

طباع، ف؛ فخرو، ع. ن؛ مسعود، و. (2024). تقييم جودة التقويم الجامعي من وجهة نظر الطلبة. ورقة بحثية في الملتقى الدولي *هندسة وتكنولوجيا التعليم، إعادة هندسة مستقبل التعليم: رؤية للحياة الطيبة والمرونة"*، جامعة قطر، 25-27 فبراير.

علام، ص. م. (2004). *التقويم التربوي البديل: أسسه النظرية والمنهجية وتطبيقاته الميدانية*. القاهرة: دار الفكر العربي.

Baartman, L. k. (2008). *Assessing the assessment: Development and use of quality criteria for competence assessment programmes*. Doctorat thesis: Utrecht University. Available online at: https://dspace.library.uu.nl/handle/1874/27155.

Baartman, L. K., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C.P. (2006). The wheel of competency assessment: Presenting quality criteria for competency assessment programs. *Studies in Educational Evaluation, 32*, 153-177.

Baartman, L.K., Bastiaens, T.J., Kirschner, P. A. Van der Vleuten, C. P. (2007). Teachers' opinions on quality criteria for competency assessment programs. *Teaching and Teacher Education*, 23, 857–867.

Baartman, L. K., Prins, F. J., Kirschner, P. A., & van der Vleuten, C. P. (2011). Self-evaluation of assessment programs: A cross-case analysis. *Evaluation and Program Planning,* 34, 206-2016.

Borghouts, L. B., Slingerland, M., & Haerens, L. (2017). Assessment quality and practices in secondary PE in the Netherlands. *Physical Education and Sport Pedagogy*, 22(5), 473-489.

Bronkhorst, L. H., Baartman, L. K., & Stokking, K. M. (2012). The explication of quality standards in self-evaluation. *Assessment in Education: Principles, Policy & Practice, 19* (3), 357–378.

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York: Routeldge.

Dierick, S., & Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation*, 27, 307–329.

Dochy, F. (2009). The edumetric quality of new modes of assessment: Some issues and prospects. In G. Joughin, *Assessment, learning and judgment in higher education* (pp. 85-114). Australia: Springer.

Gerritsen-van Leeuwenkamp, K. J., Joosten-ten Brinke, D., & Kester, L. (2017). Assessment quality in tertiary education: An integrative literature review. *Studies in Educational Evaluation,* 55, 94-116.

Gerritsen-van Leeuwenkamp, K. J., Joosten-ten Brinke, D., & Kester, L. (2018). Developing questionnaires to measure students' expectations and perceptions of assessment quality. *Cogent Education, 5* (1), 1-16.

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment.* London: The Falmer Lane.

Gulikers, J., Biemans, H., & Mulder, M. (2009). Developer, teacher, student and employer evaluations of competence-based assessment quality. *Studies in Educational Evaluation*, 35, 110-119.

Johnson, R. L., Penny, J. A., & Gordan, B. (2009). *Assessing performance: Designing, scoring, and validating performance tasks*. New York: Guilford Press.

Linn, R. L., Baker, J., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educatioal Researcher, 20 (8)*, 15-21.

McMillan, J. H. (1999). Establishing high quality classroom assessments. *Reports Descriptive (141)*, Virginia Commonwealth University.

McMillan, J. H. (2018). *Classroom Assessment Principles and Practice that Enhance Student Learning and Motivation (7th Ed)*. Hudson Street: Pearson Education.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50 (9), 741-749.

Plake, B. S., Impara, J. S. & Buckendahl, C. W. (2004). Technical quality criteria for evaluating district assessment portfolios used in the Nebraska STARS. *Educational Measurement: Issues and Practice*, 23(2), 12-16.

Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences.* Bruxelles: De Boeck.

Shah, S., McLaughlin, J. E., Eckel, S. F., Mangun, J., & Hawes, E. (2016). Evaluating the quality of competency assessment in pharmacy: A framework for workplace learning. *Pharmacy, 4* (4), 1-10.

Stiggins, R. J., Arter, J. A., Chappuis, J., & S. Chappuis, S. (2014). *Classroom Assessment for Student Learning (2nd Ed.)*. New York, NY: Pearson.

Tillema, H., Leenknecht, M., & Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning —A review of research studies. *Studies in Educational Evaluation*, 37, 25—34.