تقييم التقييم: نحو رؤية جديدة لمعايير جودة برامج تقييم الكفاءات

Assessing Assessment: Towards a New Vision of Quality Criteria for Competency-Based Assessment Programs

L'évaluation de l'évaluation : Vers une nouvelle vision de critères de qualité pour les programmes d'évaluation basés sur les compétences

فاروق طباع *

جامعة مولود معمري تيزي وزو – الجزائر

تاريخ الإرسال: 2019/01/22 تاريخ القبول: 2019/04/06 تاريخ النشر: 2019/06/22

ملخص: عرفت العديد من البلدان إصلاحات تربوية عميقة بالانتقال من التعليم التقليدي إلى التعليم بالكفاءات بإحداث تغييرات واسعة في مناهجها الدراسية وفي منظورات التقييم. ونظرا لازدياد استخدام أساليب التقييم الحديثة في مختلف الأطوار التعليمية برزت الحاجة إلى تطوير معايير جودة جديدة تتلاءم مع متطلبات تقييم الكفاءات لأن معايير الجودة الكلاسيكية للصدق والثبات غير كافية وغير قابلة للتطبيق مباشرة في برامج تقييم الكفاءات. فبعد أن اتجهت ممارسات التقييم نحو إدماج التعلم بالتقييم، واستخدام مهام واقعية معقدة، وإدماج أساليب متعددة برزت الحاجة إلى تطوير معايير جديدة ملائمة. في هذا المقال قدّم الباحث إطارا مفاهيميا شاملا لمعايير جودة برامج تقييم الكفاءات من خلال تحليل أدبيات التقييم الحديثة التي انتقلت من ثقافة العملية الاختبارية إلى ثقافة التقييم، ومن معايير الجودة الكلاسيكية إلى معايير الجودة الحديثة التي تُقيّم على أساسها التقييمات الحديثة.

الكلمات المفتاحية: التقييم؛ جودة التقييم؛ معايير الجودة؛ برامج تقييم الكفاءات.

Abstract: Many countries have witnessed profound educational reforms by moving from traditional to competency-oriented education and making broad changes in their curricula and assessment perspectives. As the use of new assessment methods in different educational levels increased, the need to develop new quality criteria to cater for competency assessment requirements has emerged in response to classical quality criteria of validity and reliability that are inadequate and inapplicable to be directly adapted to competency assessment programs. After assessment practices have been renewed by integrating learning with assessment, focusing on real-life complex tasks, and integrating multiple methods; the prerequisite to develop new appropriate quality criteria has become apparent. In this paper, a comprehensive quality criteria framework is presented for competency-based assessment programs by means of analyzing the assessment literature that shifted from testing to the assessment culture and from classical quality criteria to the new quality criteria on which more recent programs will be assessed.

Keywords: Assessment; Assessment Quality; Quality Criteria; Competency-Based Assessment Programs.

Résumé : De nombreux pays, ont connu de profondes réformes éducatives en passant de l'enseignement traditionnel à l'enseignement basé sur les compétences en apportant des modifications à leurs curricula et aux paradigmes d'évaluation. Compte tenu de l'utilisation croissante de nouvelles méthodes d'évaluation dans les différentes phases de l'enseignement, de nouveaux critères de qualité adaptés sont apparus, vu que les critères de qualité classiques de validité et de fiabilité se sont montrés insuffisantes et inapplicables aux programmes d'évaluation des compétences. L'évaluation a engendré un changement dans les pratiques en intégrant l'apprentissage à l'évaluation, en recourant aux tâches complexes, et à l'intégration de plusieurs méthodes, chose qui nécessite de nouveaux critères appropriés. Nous avons présenté dans cet article un cadre conceptuel pour les critères de qualité des programmes d'évaluation des compétences à travers l'analyse de la littérature, qui a marqué le passage de la culture de testing à la culture de l'évaluation, et des critères classiques aux nouveaux critères sur lesquelles reposent les nouvelles évaluations.

Mots clés : évaluation ; qualité d'évaluation ; critères de qualité ; programmes d'évaluation des compétences.

* المؤلف المراسل: ftebbaa05@yahoo.fr

مقدمة:

في السنوات الأخيرة اتجهت العديد من بلدان العالم إلى إجراء إصلاحات تربوية عميقة للاستجابة للتحديات التي فرضتها التطورات التكنولوجية ونظم المعلومات ومتطلبات سوق العمل بالاعتماد على التعليم القائم على الكفاءات الذي يسعى إلى إعداد أفراد قادرين على مواجهة وضعيات الحياة المهنية والاجتماعية المعقدة. وقد جاءت هذه الاصلاحات التربوية لتضييق الفجوة بين ما يتم تعلّمه في المدرسة ومتطلبات سوق العمل الذي يفرض باستمرار تطوير كفاءات جديدة للمهن الحالية وتطوير مهارات التعلم مدى الحياة (Baartman, Gulikers, & Dijkstra, 2013).

ولتدعيم الاصلاحات التربوية نادى العديد من المختصين والباحثين إلى تطوير أساليب تقييم أخرى قادرة على البرهنة والاستدلال بشكل كافي على كفاءات الطلاب، وأصبح يعتبر الاعتماد على الأساليب الحديثة للتقييم حجر الزاوية في تنفيذ التعليم بالكفاءات. وذلك على اعتبار أنه إذا أُريد للإصلاحات التربوية نجاحها يجب أن تعطي للتقييم أولوية ومكانة هامة، ويجب أن تركز مقاربة التقييم على إدماج المعارف والمهارات والاتجاهات (Baartman, 2008).

من بين أساليب التقييم الحديثة التي طوّرت للاستدلال على كفاءات الطلاب؛ ملفات الانجاز، والمشاريع، وعينات الكتابة، والعروض الشفهية والكتابية، والملاحظة، والاختبارات الشاملة، والمحاكاة، والاجابات المبنية، والتقييم الذاتي، وتقييم الأقران، وغيرها من أساليب التقييم (Birenbaum & Dochy, 1996; Dierick & Dochy, 2001; Scallon, 2004; Segers et al., 2003؛ علام، 2004).

وتجدر الإشارة إلى أن أساليب التقييم الحديثة لا تلغي استخدام الأساليب التقليدية وإنما تكون مندمجة ومتكاملة ضمن ما يسمى ببرامج تقييم الكفاءات، وإدماج أساليب تقييم متعددة في برامج تقييم الكفاءات ليست فكرة قديمة وإنما فكرة جديدة ظهرت بسبب عدم كفاية استخدام طريقة تقييم واحدة في برامج تقييم الكفاءات، كما أن أساليب التقييم الحديثة والتقليدية المختلفة تدمج مع بعضها بشكل متكامل للتحقق من الأهداف التربوية، وبهذا فهي تشكّل برنامج التقييم يتم تطبيقه في الممارسات التعليمية.

وتهتم برامج تقييم الكفاءات بالأساس بتقدير الإنجازات الحقيقية للطلاب على توظيف المعارف والمهارات بطريقة إبداعية في حل المشكلات (Birenbaum, 1996)، وقد ظهرت نتيجة للآثار والعواقب السلبية للتقييمات التقليدية (عبارات الاختيار من متعدد، والصحيح والخطأ، والربط...وغيرها) التي تهتم بالأساس بفحص معارف الطلاب الأمر الذي فرض الانتقال إلى استخدام مهام تقييم مستمدة من مشكلات الحياة الواقعية أو صور حقيقية من مشكلات الحياة الواقعية للطالب (Segers, Dochy, & Cascallar, 2003)، وبما أن هدف التربية حاليا هو تعليم الطلاب كيفية استخدام معارفهم لحل المشكلات الواقعية فان أساليب التقييم الحديثة تبقى أدلة مباشرة وصادقة على اكتساب الكفاءات مقارنة بالاختبارات المقننة (Dochy, 2009).

تتميز برامج تقييم الكفاءات بمميّزات وخصائص متعددة فريدة من نوعها كإدماج التعلم بالتقييم، وقياس عمليات التفكير العليا، واستخدام مهام معقدة، والأداء في وضعيات واقعية، واستخدام تصحيح محكي، وإدماج أساليب تقييم

متعددة (Baartman, Bastiaens, Kirschner, & van der Vleuten, 2006; Birnbaum, 1996 ; Dochy, 2001 ; 2009 ; Dierick & Dochy, 2001; Gulikers, Biemans & Mulder, 2009 ; Scallon, 2004; Segers et al., 2003).

وفي السنوات الأخيرة اهتم بعض المؤلفين بتقييم جودة التقييم أو ما يسمى لدى بعض المؤلفين بـ "تقييم التقييم" (Baartman, 2008 ; Linn, Baker, & Dunbar, 1991) نظرا لأهميتها في برامج تقييم الكفاءات لأن القرارات المتخذة على الطلاب في مختلف المستويات الدراسية تعتمد على نواتج هذه البرامج، لذا لا يتم تجاهل تقييم جودة برامج تقييم الكفاءات. ومن الأسباب التي أدت إلى التوجه إلى تقييم جودة التقييم هي ثقافة التقييم التي ترى بضرورة إدماج التعلم بالتقييم والتدريس، ومشاركة الطلاب في عملية التعلم، وواقعية ودلالة مهام التقييم، والتركيز على كلا من العمليات والنواتج في التعلم، والانتقال إلى التقييم الوصفي القائم على القدرات والنواتج (Segers et al., 2003).

ونظرا لازدياد استخدام برامج تقييم الكفاءات في مختلف المراحل التعليمية برزت الحاجة إلى تطوير معايير جودة جديدة وملائمة، وذلك لأن معايير جودة الاختبارات الكلاسيكية (مثلا، الصدق والثبات) لا يمكن تطبيقها كما هي في برامج تقييم الكفاءات (Baartman, 2008; Baartman et al., 2006 ; Dochy, 2009) لأنّه من أجل الحصول على تقييمات ذات جودة عالية من الضروري تقديم معايير جودة جديدة وملائمة كالأصالة، والتعقيد المعرفي، والدلالة، وإمكانية إعادة إنتاج القرارات، والعواقب التربوية، والتكاليف والفعالية.

بعض معايير جودة التقييم نشأت من ثقافة العملية الاختبارية التي استمدت من المقاربة السيكومترية كالصدق والثبات والانصاف والشفافية، وبعض المعايير نشأت من ثقافة التقييم التي استمدت من أفكار المقاربة الايديومترية كالأصالة والدلالة والعواقب التربوية. وبسبب إدماج برامج تقييم الكفاءات كلا من أساليب التقييم الحديثة وأساليب التقييم التقليدية في برامج تقييم الكفاءات فمن الضروري أن تتكامل وتندمج كلا من معايير الجودة الكلاسيكية ومعايير الجودة الجديدة(Baartman, 2008; Baartman et al., 2006 ; Baartman et al., 2007 ; Baartman, Prins et al., 2007).

وفي ظل التطورات في أساليب التقييم الحديثة التي نشأت من ثقافة التقييم أسست بدورها لمعايير الجودة الجديدة، وفي هذا المقال تم تناول معايير جودة برامج تقييم الكفاءات بتحليل مختلف التطورات التي حدثت بالانتقال من ثقافة العملية الاختبارية إلى ثقافة التقييم، والتعريف ببرامج تقييم الكفاءات، وتقديم معايير الجودة الكلاسيكية (الصدق والثبات) ومعايير الجودة الجديدة وصولا إلى تقديم إطار مفاهيمي شامل لمعايير جودة برامج تقييم الكفاءات.

1-الانتقال من ثقافة العملية الاختبارية إلى ثقافة التقييم:

خلال العقدين الأخيرين حدثت تغييرات هامة في التقييم التربوي، والتي نتجت من الانتقال من ثقافة العملية الاختبارية إلى ثقافة التقييم (Birenbaum, 1996; Dierick & Dochy, 2001; Gipps, 1994; Segers et al., 2003)، واستمدت ثقافة العملية الاختبارية أسسها ومبادئها من النظرية السلوكية في التعلم والتدريس التي تعتبر المتعلم متلقي سلبي للمعرفة، وتتوجه بالدرجة الأولى أثناء عملية التقييم إلى قياس المعارف والمهارات الدنيا للمتعلمين (Birenbaum, 1996;

(2003. فالتقييم والتدريس منفصلان عن بعضهما البعض أين يقوم المدرس بالتدريس وخبراء قياس خارجيون يتكفلون بتطوير أدوات تقييم تستخدم من طرف المدرس، كما أن التقييم منفصل عن عملية التعلم أين تختلف فيها فترات التعلم عن فترات التقييم.

أما أساليب التقييم أكثر شيوعا واستخداما في ثقافة العملية الاختبارية هي أسئلة الاختيار من متعدد والصحيح وأسئلة الخطأ والربط التي تطبق في ظروف صارمة أين لا يُسمح للمتعلم باستخدام الوسائل والأدوات المساعدة، كما تهتم بتقييم النواتج النهائية دون الاهتمام بالعمليات التي تؤدي إلى النواتج (Baartman et al., 2007)، كما أن عملية تطوير الاختبارات والمعايير لاصدار الأحكام غير معروفة لدى الطلاب، وتعتمد جودة التقييم على النظرية السيكومترية في تطوير وتصحيح وتفسير نتائج الاختبارات كما تُقيّم على بالأساس وفق معياري الصدق والثبات (Dierick & Dochy, 2001)).

أما ثقافة التقييم فقد نشأت من تزايد الانتقادات الموجهة لأساليب العملية الاختبارية التقليدية التي تتميّز بالطبيعة غير الواقعية لاختباراتها، وضعف الثقة فيها كقياسات صادقة للتعلم، وزيادة اعتمادها على الاختبارات كهدف نهائي لعملية التدريس. وقد استمدت ثقافة التقييم أسسها من النظرية المعرفية التي تنظر إلى التعلم بأنه عملية بنائية نشطة وتجربة مثيرة تحدث في سياق واقعي وتؤدي إلى تعلم ذات معنى (Birenbaum, 1996). وتستخدم في الاستدلال على كفاءة الطلاب على أساليب تقييم حديثة أقل تقنينا من الأساليب التقليدية المستمدة من وضعيات واقعية أصيلة ومحفزة وذات معنى، ومن بين هذه الأساليب ملفات الانجاز، والاختبارات الشاملة، والتقييم الذاتي التي تستخدم متكاملة مع الأساليب التقليدية. كما يهدف التقييم إلى تحقيق أغراض تكوينية وتجميعية معا بعناية أين يتميّز بالاستمرارية من أجل الضبط وتقديم تغذية راجعة فورية، ويُقيّم المدرس في التعلم كلا من النواتج والعمليات، ويوثّق الطالب تقدّمه على أساس ملفات الانجاز مثلا (Birenbaum, 2003).

يعتبر المتعلم في ثقافة التقييم عنصرا مشاركا في التعلم من خلال التعاون مع المدرس وزملائه، وفي تقييم ذاته وتقييم أعمال زملائه، ومشاركا في وضع معايير التصحيح (Dierick & Dochy, 2001). وعلى اعتبار أن الكفاءات سمات غير ثابتة ومتطوّرة بمرور الوقت فإنها تحتاج إلى نظام جديد لتقييم جودتها مما أدى إلى تأسيس مدرسة للاديومترية عوضا عن السيكومترية (Dochy, 2001; Dochy, 2009 ; Dierick & Dochy, 2001)، وعليه تُقيم جودة برامج تقييم الكفاءات على أساس معايير جديدة أخرى مستمدة من المقاربة الايديومترية (Dochy, 2009).

قبل أن نتناول معايير الجودة الكلاسيكية (الصدق والثبات) ومعايير جودة برامج تقييم الكفاءات، نحاول أن نقدم وصف مختصر لبرامج تقييم الكفاءات التي تتطلب رؤية جديدة لمعايير جودتها تتلاءم مع خصائصها الفريدة من نوعها.

2-برامج تقييم الكفاءات:

في إطار ثقافة التقييم طوّرت أساليب تقييم جديدة كالإجابات المبنية، وملفات الانجاز، والمشاريع، والملاحظات، والاختبارات الشاملة، والتقييمات الذاتية، وتقييمات الأقران (Birenbaum & Dochy, 1996; Dierick & Dochy,

2001; Feuer & Fulton, 1993 ; Scallon, 2004; Segers et al., 2003، علام، 2004). وتستخدم أساليب التقييم للبرهنة على كفاءات الطلاب كعملية معقدة ومركبة لإدماج المعارف والمهارات والاتجاهات (Scallon, 2004)، وبالتالي فمن المستحيل الاستدلال على امتلاك الطالب للكفاءة المطلوبة باستخدام أسلوب تقييم واحد (Baartman et al., 2006).

أكدت بارتمان وآخرون (Baartman et al., 2006) على ضرورة إدماج أساليب التقييم الحديثة مع أساليب التقييم التقليدية في برامج الكفاءات لأنّ تقييم الكفاءات ليس مشكلة سيكومترية يتطلب حلّها بواسطة أسلوب تقييم واحد، ولكنها مشكلة تصميم تعليمية تتضمن كل أساليب التقييم في المنهاج (Van Der Vleuten & Schuwirth, 2005).

وعلى الرغم من تطوير أساليب تقييم حديثة التي ينتظر منها مستقبل واعد إلا أن الاختبارات التقليدية لا يجب استبعادها في برامج تقييم الكفاءات. تؤكد هذه الفكرة على أن أساليب التقييم الحديثة ليست أساليب بديلة للاختبارات التقليدية وإنما مكمّلة تسمح بأن تؤدي أدوارا مشتركة رغم أنهما نابعتين من ثقافتين متناقضتين (Birenbaum, 1996)، وفي هذا الإطار تؤكد بارتمان وآخرون (Baartman et al., 2007) أن أساليب التقييم الحديثة ليست بديلة للاختبارات التقليدية لأنها عناصر ضرورية يمكن أن تفيد في تحقيق أهداف معينة، لذا فمن الضروري أن تستخدم متكاملة ومندمجة مع الأساليب الحديثة (Dochy, 2009)

عرفت بارتمان وآخرون (Baartman et al., 2007) برامج تقييم الكفاءة بأنها مزيج من الأساليب التقليدية والأساليب الحديثة، ويتوقف مزجها على أهداف البرامج التربوية والكفاءات المراد تقييمها وتوسع البرنامج، مع ضرورة أن يغطي برنامج تقييم الكفاءات كل الأهداف التربوية، وضرورة تقييم المعارف والمهارات والاتجاهات بطريقة مندمجة.

يتم في برامج التقييم الحالية الجمع بين أساليب التقييم المختلفة بما فيها اختبارات تقييم المعارف كالاختيار من متعدد أو الأسئلة المفتوحة، والأساليب المطوّرة حديثا كملفات الانجاز والمقابلات المعتمدة على المعايير وتقييمات الأداء، ويمكن أن يكون للتقييمات وظائف تكوينية وتجميعية حسب سياق وأهداف البرنامج التعليمي (Bronkhorst, Baartman, & Stokking, 2012). فعلى سبيل المثال قام الباحثون بارتمان وبرينز وآخرون (Baartman, Prins et al., 2007) بتقديم برنامج تقييم الكفاءات بواسطة المشروع كأسلوب تضمّن مزيج من الاختبارات (الاختيار من متعدد، والأسئلة المفتوحة) وتقييم المنتوج بواسطة المشاريع، والمقابلة، والملاحظة داخل وخارج القسم.

3–معايير جودة برامج تقييم الكفاءات:

إن ضمان جودة برامج تقييم الكفاءات عملية مهمة جدا لا يمكن تجاهلها لأن القرارات المتخذة حول الطلاب تستند إلى تلك البرامج، وتتوقف جودة الاختبارات التقليدية على معياري الصدق والثبات، ولكن هل هما كافيان لتحديد جودة برامج تقييم الكفاءات؟ وهل تتطلب معايير أخرى جديدة أكثر شمولية واتساعا وتكيّفا مع أساليب التقييم الحديثة؟

3-1-معايير الجودة الكلاسيكية:

يشير الصدق إلى مدى قياس أداة التقييم لما وضعت لقياسه ومدى ارتباط التقييم بالهدف، أما الثبات فيشير إلى مدى اتساق نتائج التقييم الذي يستدل عليه بالحصول على نفس النتائج إذا قدّرت بواسطة شخص آخر أو في فترة أخرى.

وبالأخذ بعين الاعتبار الخصائص المتميّزة لأساليب التقييم الجديدة ناقش العديد من المؤلفين مسألة عدم كفاية معايير الجودة الكلاسيكية للصدق والثبات (Baartman et al., 2006; Baartman, Prins et al., 2007; Dochy, 2009; Scallon, 2004; De Ketele & Gerard, 2005؛ طباع، 2017). وقد وصف كرونباخ ولين وبرينان وهيرتل (Cronbach, Linn, Brennan, & Haertel, 1997) السمات الأساسية البارزة لأساليب التقييم الجديدة التي جعلت من الطرق الكلاسيكية لتحليل خطأ القياس غير ملائمة، وهي أنها تقييمات محكية المرجع، والمهام المستخدمة ذات نهاية مفتوحة ومعقدة. لذا يتطلب في الوقت الحالي البحث عن قياسات جودة أخرى جديدة لاصدار أحكام بطريقة مسؤولة ومنصفة (Gipps, 1994).

إن استخدام معايير الجودة التقليدية للثبات والصدق تحوم حولها شكوك في قدرتها على معالجة مسائل إمكانية التعميم، وصدق الاستدلالات مقارنة التقييمات الجديدة. لذا فمن المهم توسيع المعايير المستخدمة للحكم على مدى كفاية تقييم أساليب التقييم الحديثة التي نستخدمها بشكل واسع في الممارسات التعليمية (Linn et al., 1991). وفي هذا الإطار أوضح الباحثون بارتمان وبرينز وآخرون (Baartman, Prins et al., 2007) وجود مشكلات في استخدام مفهومي الصدق والثبات في تقييم جودة برامج تقييم الكفاءات، وقدموا طرقا حول كيفية تكييفها حتى تتلاءم مع أساليب التقييم الحديثة.

كما يمكن أن تنجم عن استخدام فكرة الثبات في مجال تقييم الكفاءات بعض المشكلات (van der Vleuten & Schuwirth, 2005)، فالطرق الكلاسيكية المستخدمة لتقدير الثبات في الاختبارات التقليدية مهمة، ولكنها غير ملائمة في مراقبة وفحص جودة برامج تقييم الكفاءات (Baartman et al., 2006). وبالتالي من الضروري تقديم طرق جديدة في تقدير ثبات أساليب تقييم الكفاءات تختلف عن طرق الثبات في الاختبارات الكلاسيكية.

أما فيما يتعلق بالصدق فان القليل من المؤلفين الذين تناولوا هذه المسألة، والأدلة التي تتوفر في الوقت الحالي استوحيت من الطرق الكلاسيكية، حيث يخضع الصدق إلى مكونين أساسيين، وهما: مهام التقييم وأساليب إصدار الحكم، لذا فأدوات تقييم الكفاءات تعاني من الغموض والضبابية (Scallon, 2004).

كما تكمن المشكلة الأساسية للصدق في تعريفاته المتعددة التي وردت في أدبيات القياس، وقد كانت مساهمات ميسيك (1994; 1995) Messick ذات أهمية بتقديم نموذج موحد ومتكامل للصدق يميّز ستة مصادر لأدلة صدق التكوين الفرضي، وهي: أدلة تعتمد على المحتوى، وأدلة تعتمد على عمليات الاستجابة والتأصيل، وأدلة تعتمد على البنية الداخلية، وأدلة تعتمد على العلاقات بمتغيرات أخرى، وأدلة تعتمد على إمكانية التعميم، وأدلة تعتمد على عواقب العملية الاختبارية. وقد تضمن الاطار الذي قدمه "ميسيك" أدلة صدق عواقب استخدام وتفسير نتائج التقييم الذي اعترفت به الجمعية

الأمريكية للبحث التربوي والجمعية الأمريكية لعلم النفس والمجلس القومي للقياس (American Psychological Association, Educational Research Association, & National Council on Measurement in Education, 2014)، وفي نفس الاطار قام ميلر ولين (Miller & Linn, 2000) بتحليل أدبيات البحوث والدراسات التي أجريت لتأييد صدق أساليب التقييم الجديدة (تقييمات الأداء) من خلال عرض النتائج التي توصلت إليها في كل مصدر من مصادر أدلة الصدق التي قدّمها "ميسيك".

اتساع مفهوم الصدق وتعقيده مشكلة أخرى يمكن أن تؤدي إلى صعوبة في فصل أبعاده المتداخلة في الممارسة العملية، لذا هناك حاجة ملحة إلى مقاربة جديدة لإيجاد حلول لمسائل تأييد الصدق، وهنا قدم كاين (Kane, 2001) مقاربة قائمة على الحجج، وهي مهمة لجميع مصادر أدلة الصدق لكنها لا تعطى تعريف واضح للصدق للممارسين حول الأدلة المطلوب جمعها من أدوات التقييم المستخدمة (Baartman, Prins et al., 2007). ومن هنا أصبحت مسألة توضيح مفهوم الصدق وزيادة تفعيله للاستخدام الفعلي والصحيح مع أساليب التقييم الحديثة ضرورة ملحة في الممارسة العملية.

في إطار برامج تقييم الكفاءات تغيّرت النظرة إلى المعايير الكلاسيكية للصدق والثبات من الناحية النظرية والتطبيقية، وقد أدى هذا إلى تقديم معايير جودة جديدة ملائمة تعوّض مفهومي الصدق والثبات، ووسّعت من معايير الجودة. وقد قدم في هذا الاطار لين وآخرون (Linn et al., 1991) إطارا شاملا ومتكاملا لمعايير الجودة ذات أهمية في تأييد صدق تقييمات الكفاءة. فعلى سبيل المثال الثبات كمفهوم كلاسيكي يتعلق باتساق نتائج التقييم عبر المهام أو الفترات أو المقيّمين، ولكن من منظور أساليب التقييم الحديثة تحوّل الثبات إلى "إمكانية التعميم" عبر عدة تقييمات، وعدة مقيمين وعدة فترات، وأدرج الثبات ضمن معيار "إعادة إنتاج القرارات" عبر المقيّمين، والفترات، والسياقات، وأساليب التقييم المستخدمة (Baartman, 2008 ; Tilema et al., 2011)، كما يجب أن تجرى التقييمات في سياقات أقل تقنينا، وتستند إلى مهام متسقة ومتجانسة حتى وإن لم تكن متطابقة، ويندرج الثبات وفقها ضمن معيار "إمكانية المقارنة" (Baartman, 2008). من هنا أصبح الثبات يتعلق بالحصول على نتائج قابلة للتعميم عبر نطاقات أوسع من المهام والمقدرين والفترات، فلا يكتفي صانع القرار بالاتساق عبر بُعد واحد وإنما بإمكانية مقارنة النتائج عبر أبعاد تقييم مختلفة لتحقيق درجة عالية من الموثوقية.

يتعلق الصدق كمفهوم كلاسيكي بمدى قياس أداة التقييم لما أعدت لقياسه، ولكن أصبح الصدق حاليا حكم تقييمي شامل لمدى قدرة الأدلة الامبريقية ومنطق التأصيل النظري على تدعيم كفاية وملاءمة التفسيرات والقرارات على درجات المقاييس أو نماذج أخرى من القياس لتلبية الغرض (Messick, 1995). فعلى سبيل المثال انتقل معيار الصدق من الاقتصار على تلاؤم التقييم مع الهدف المطلوب "الملاءمة مع الهدف" إلى الاهتمام بمعايير أكثر واقعية مثل "التعقيد المعرفي" و"الأصالة" و"العواقب التربوية"، فعلى أساس معايير الجودة الجديدة المذكورة (ومعايير جودة أخرى إضافية) التي سوف نقدّمها بنوع من التفصيل يقدم مفهوم الصدق الحديث أدلة كافية عن ملاءمة وفعالية وجودة برامج تقييم الكفاءات.

وبالتالي فان مراقبة جودة برامج تقييم الكفاءات على أساس معايير ظهرت في فترة سابقة غير ممكنة، وأصبحت الطرق الكلاسيكية للصدق والثبات محل شكوك لإمكانية إعطائها صورة خاطئة عن النتائج، إضافة إلى أن استخدام

أساليب التقييم الحديثة يعنى بالضرورة تقدير المقيمين للمعارف والمهارات بطرق متعددة وفي أوقات مختلفة، فتحقيق أساليب تقييم منصفة ومتميزة تحتاج المعايير التقليدية إلى توسيعها أكثر فأكثر، إضافة إلى تقديم معايير جودة أخرى أكثر ملائمة.

3-2-معايير الجودة الحديثة:

بالإضافة إلى توسيع المفاهيم الكلاسيكية للصدق والثبات في إطار تقييم الكفاءات، اقترحت معايير جودة جديدة واسعة ومتكاملة مستمدة من ثقافة التقييم (Dierick & Dochy, 2001; Linn et al., 1991). وأكد في هذا الشأن الباحثون الأوائل على غرار لين وزملائه (Linn et al., 1991) على ضرورة توسيع فكرة الجودة حتى تتلاءم مع خصائص أساليب التقييم الحديثة، وقدموا معايير جديدة، وهي: العواقب، والانتقال وإمكانية التعميم، والانصاف، والتعقيد المعرفي، وذات المعنى، وجودة المحتوى، وتغطية المحتوى، والتكلفة والفعالية. وقد ذكر جونسون وآخرون2009 ,.Johnson et al)) معايير جديدة تتمثل في الأصالة، والسياق، والتعقيد المعرفي، والتغطية العميقة للمحتوى، والإجابة المبنية من المفحوص، والمصداقية، والتكاليف، والاصلاح.

كما قدم دوتشي (Dochy, 2009) معايير جودة متضمنة التعقيد المعرفي، والأصالة، والوجاهة، والانصاف، والشفافية، وقدّمت بارتمان وآخرون (Baartman et al., 2006) معايير جودة أكثر تحديدا وشمولية وردت في الإطار الفكري لمعايير جودة برامج تقييم الكفاءات الذي اشتمل على عشرة معايير، وهي: الأصالة، والتعقيد المعرفي، وذات المعنى، والانصاف، والشفافية، والعواقب التربوية، والوجاهة، وإمكانية إعادة إنتاج القرارات، وإمكانية المقارنة، والتكاليف والفعالية.

معايير جودة برامج تقييم الكفاءات المقدمة في هذا المقال (أنظر إلى الجدول 1) استندت إلى العديد من الأدبيات المتخصصة في هذا المجال (Baartman, 2008; Baartman et al., 2006; Baartman et al., 2007; Dierick & Dochy, 2001; Dochy, 2009; Gulikers, Bastiaens & Kirschner, 2004; Johnson et al., 2009; Linn et al., 1991; Tillema et al., 2011)، واستمدت بالأساس من الاطار الفكري لمعايير جودة برامج تقييم الكفاءات الذي قدمته بارتمان وآخرون((Baartman, et al., 2006; Baartman, Prins, et al., 2007).

ونظرا لشهرة هذا الاطار الفكري والصدى الواسع الذي عرفه في أدبيات تقييم جودة برامج التقييم سواء في التعليم العام أو التعليم المهني أو التعليم الجامعي، قام العديد من الباحثين باستخدامه في دراسات تقييم جودة التقييمات في العديد من المجالات والسياقات (Baartman, Prins, Kirschner, & van der Vleuten, 2011; Bronkhorst et al., 2012; Gulikers, Biemans, & Mulder, 2009; Jonsson, Baartman, & Lennung, 2009; Shah, McLaughlin, Eckel, Mangun, & Hawes, 2016)، وبالتركيز على المفاهيم والتعريفات التي استخدمها المؤلفون حدّدنا بدقة واختصار معايير الجودة في إطار شامل ومتكامل مع تفادي تكرار معايير الجودة التي تحمل معاني متقاربة، وتم تجميعها في عناوين عريضة.

1. التلاؤم مع الهدف: يتعلق بتوافق أهداف التربية مع أهداف التقييم بشكل أكثر تلاحما، وضرورة الانسجام بين المعايير والمنهاج والتقييم وأهداف التقييم والطرق المستخدمة في التقييم مع أهداف التربية (Baartman, Prins et al.,

2011)، وضرورة أن تتوافق أهداف المنهاج مع أهداف وأساليب ومعايير ومحكات وأساليب تقييم الكفاءة (Miller & Linn, 2000).

2. التلاؤم مع التقييم الذاتي: يجب أن تحفز برامج تقييم الكفاءة استراتيجيات التعلم المنظم ذاتيا، والذي أصبح هدفا مهما وبارزا للتعلم، ويجب أن تتضمن أساليب معينة تعزّز بعض التعلّمات كالتقييم الذاتي، وتقديم وتلقي التغذية الراجعة (Baartman, Prins, et al., 2007; Baartman et al., 2011)، ويمكن أن تحفز برامج تقييم الكفاءة استراتيجيات التعلم المنظم ذاتيا عن طريق التقييم الذاتي، وإعطاء حرية للطلاب في وضع أهداف تعلمهم بأنفسهم.

3. الأصالة: معيار مهم يشير إلى أن يعكس التقييم إمكانية تطبيق المعارف والمهارات الأساسية في الحياة الواقعية والمهنية (Johnson et al., 2009) أو أن يعكس تشابه برنامج التقييم مع الحياة المهنية للطلاب، وضرورة تقييم الكفاءات المطلوبة في العمل المستقبلي، وتتحدد الأصالة في أربعة جوانب: مهمة التقييم، والسياق الطبيعي والاجتماعي، وأسلوب التقييم أو نتيجته، ومعايير التقييم (Gulikers et al., 2004). بالإضافة إلى تقديم مهام واقعية لقياس المعارف والمهارات والاتجاهات بشكل مندمج ومتكامل في سياق واقعي (Dierick & Dochy, 2001).

4. التعقيد المعرفي: يشبه التعقيد المعرفي معيار الأصالة لأنه يتعلق أيضا بالحياة المهنية المستقبلية، ويركز بشكل مباشر على ضرورة أن تعكس مهام التقييم استدعاء المهارات المعرفية العليا (Linn et al., 1991). ويعد من أكثر المميزات أهمية في تقييم الكفاءات لأنه يقوم على قياس الاستراتيجيات المعرفية العليا كبناء مشكلة أو مهمة، وإعداد خطة لمعالجة مشكلة، وتوظيف المعلومات، وبناء إجابات، وتوضيح أساليب تطوير الاجابة (Johnson et al, 2009). وتجدر الاشارة أنه يجب أن تكون مهام التقييم ذات تعقيد معرفي كافي لاستثارة عمليات التفكير العليا وحل المشكلات التي يصادفها الطالب في المجال المهني.

5. الانصاف: ينص الانصاف على أن برنامج تقييم الكفاءات لا يظهر تحيزا لمجموعات معينة من الطلاب، وإبعاد التباين غير الموائم في مهام التقييم (Linn et al., 1991; Messick, 1994)، وبتعبير أكثر دقة يعني إعطاء جميع الطلاب فرص متساوية للبرهنة على قدراتهم وكفاءاتهم وتعزيز إمكانياتهم أثناء التقييم (Dierick & Dochy, 2001))، ويمكن أن ترجع الأسباب المحتملة للتحيز والتباين غير الملائم إلى الجوانب اللغوية أو الثقافية غير المألوفة لدى بعض الطلاب في المهام.

6. ذات الدلالة: يجب أن يكون برنامج تقييم الكفاءات ذات معنى أو ذات قيمة دالة لدى الطالب والمعلم على حد سواء (Linn et al., 1991; Messick, 1994)، ولدى أرباب العمل في المستقبل، وتزداد دلالة التقييم بإشراك الطلاب في تطوير التقييم، وإدراكهم بوجود صلة بين مهام التقييم واهتماماتهم الشخصية، ويمكن أن يصبح التقييم أكثر قيمة لدى الطلاب عندما يكونوا قادرين بأنفسهم على تحديد مدى استعدادهم للتقييم وتحقيق فائدة منه (Baartman et al., 2006)، وتفرض دلالة التقييم أن يكون الطالب قادرا على ربط علاقة وجدانية ايجابية مع المهام وتثير لديه الرغبة في الانجاز (Roegiers, 2000). يتطلب المعيار أن تعطى أهمية لمحتوى مهام التقييم حتى تكون جديرة بالاهتمام ومفيدة في تقييم مجهودات الطالب.

7. الوجاهة: تشير الوجاهة إلى قابلية تفسير الأساتذة أو المقيّمين نتائج التقييم بشكل فوري دون تأويلات لأن تقييم الأداء العملي يكشف مباشرة على كفاءة الطالب في وضعية حقيقية(Baartman, et al., 2006; Baartman, Prins et al., 2007)، فالطريقة التي يتم بها تقييم الكفاءة بشكل مباشر أو تفسير إجابات الطالب له تأثير فوري على تعليمه وتعلّمه (Dochy, 2009). وقد أشار لين وآخرون (Linn et al., 1991) إلى إمكانية تعزيز الصدق بواسطة التقييمات المباشرة للأداء، وتوجد العديد من الأدلة على قدرة أساليب التقييم المباشرة على التنبؤ بالنجاح في العمل أفضل من الأساليب غير المباشرة.

8. الشفافية: تعني الشفافية مدى وضوح وإمكانية فهم برنامج تقييم الكفاءات من طرف جميع المهتمين (مدرسين، طلاب، أرباب عمل) (Baartman et al., 2006; Dierick & Dochy, 2001)، إذ يجب على الطلاب معرفة الغرض من التقييم ومعايير التصحيح والقائمين بعملية التصحيح (Baartman, 2008) وأن يعرفوا المطلوب منهم حتى يكونوا قادرين على الاعداد للتقييم وضبط تعلّمهم الذي يحقق أهدافهم، كما يجب على المعلمين والمقيّمين معرفة وفهم برنامج تقييم الكفاءات بشكل كامل حتى يتم إعدادهم للقيام بدورهم بشكل أفضل. بالإضافة إلى تأثير شفافية التقييم على عملية تعلّم الطلاب لأن أداءهم سوف يتحسّن إذا كانوا على دراية بمعايير التصحيح (Dochy, 2009).

9. العواقب التربوية: ذكرت العواقب التربويةكمعيار مهم جدا في برنامج تقييم الكفاءات من طرف العديد من المؤلفين (Dierick & Dochy, 2001; Linn, et al., 1991; Messick, 1994) والذي يتعلق بالآثار المرجوة وغير المرجوة الايجابية والسلبية المترتبة عن استخدام التقييم (Linn et al., 1991). ويمكن أن تتضمن العواقب الايجابية زيادة دافعية تعلم المفاهيم المهمة، وحل المشكلات، وتوظيف ما تعلمه الطلاب، واستفادة صانعي القرارات والمسؤولين من المعلومات لاتخاذ قرارات حول الطلاب (Johnson et al., 2009)، وتتضمن العواقب السلبية اتخاذ قرارات متحيزة، وانخفاض في مستوى الطلاب، والاتجاهات السلبية للمعلمين، وضعف التكوين، والتكاليف المرتفعة للتقييم.

10. إمكانية إعادة إنتاج القرارات: يشير إلى ضرورة استناد القرارات المتخذة حول الطلاب إلى تقييمات متعددة يقوم بها مقيمون متعددون في فترات مختلفة. والذي يسمى أيضا بانتقال وإمكانية التعميم (Linn et al., 1991) وبإمكانية التعميم والضمان (Tilema et al., 2011). وبتعبير أكثر دقة يشير إلى إمكانية تعميم نتائج التقييم على نطاق أوسع من الطلاب والمهام والمقيّمين، وأن لا يتوقف الغرض من التقييم على أداء مهمة واحدة تُقيم بمصحح واحد خلال فترة واحدة، ولكن يجب أن تُعمّم على نطاق أوسع من المهام والمصححين والفترات وأساليب التقييم(Scallon, 2004). وهذا لا يعني أن برنامج تقييم الكفاءات يجب أن يكون موضوعيا، وإنما يجب البحث على نتائج تقييمات قابلة للتعميم قدر الإمكان.

11. إمكانية المقارنة: في الواقع تشير إمكانية المقارنة إلى ضرورة أن يُجرى برنامج تقييم الكفاءة بطريقة متسقة وموثوقة (Baartman, 2008)، بمعنى أن تكون الظروف التي يُجرى فيها التقييم نفسها قدر الإمكان لكل الطلاب وتُصحح بطريقة متسقة بواسطة نفس المعايير. وتتضمن إمكانية المقارنة الزيادة في اختيار ظروف التقييم باستخدام عينة واسعة من

محتوى ومهام الكفاءة (Baartman, et al., 2006). وعلى العموم نظرا لضعف تقنين أساليب تقييم الكفاءة فإنها تتطلب أحكام المقيّمين، لذا فان اتساق إجراءات التصحيح عملية مهمة جدا (Scallon, 2004).

12. التكاليف والفعالية: معيار مهم جدا لأن برامج تقييم الكفاءات غالبا ما تكون صعبة التطبيق مقارنة بالاختبارات التقليدية بالرغم من قدرتها على تحقيق نواتج تعليمية مهمة. تشير الفعالية إلى تصميم مهام تقييم محفزة تعتمد على مقاربات تعلم عميقة كأداة فعالة لتعزيز التعلم والاهتمام وتقديم تغذية راجعة للفهم والضبط والجودة والتحفيز والاعداد للحياة (Boud & Falchikov, 2007). أما التكاليف فتشير إلى الوقت والجهود والموارد المطلوبة لتطوير وتطبيق وتصحيح التقييمات لأن تكاليف تقييمات الكفاءة في العادة تكون مرتفعة راجعة إلى نفقات استخدام المقدرين، والوقت المطلوب للتقييم (Johnson et al., 2009).

الجدول –1– : معايير جودة برامج تقييم الكفاءات

معايير الجودة	الوصف المختصر
الملاءمة مع الهدف	يجب أن يكون توافق بين أهداف المنهاج وكيفية وأساليب تقييمها، وأن تتوجه المعايير والمحكات نحو كل الكفاءات، وملاءمة الأساليب الممزوجة مع تقييم الكفاءة.
الملاءمة مع التقييم الذاتي	يجب أن تحفز برامج تقييم الكفاءات التعلم المنظم ذاتيا لدى الطلاب وصياغة أهداف تعلمهم بأنفسهم باستخدام التقييمات الذاتية.
الأصالة	مدى تشابه برامج تقييم الكفاءات مع مكان العمل في المستقبل.
التعقيد المعرفي	يجب أن تصدر برامج تقييم الكفاءات أحكام على عمليات التفكير وعلى النواتج.
الانصاف	يجب أن يحصل الطلاب على فرص عادلة لإثبات كفاءتهم بالتعبير عن أنفسهم بطرق مختلفة وعدم إظهار المقيمون للتحيز.
ذات الدلالة	يجب تكون برامج تقييم الكفاءات فرص تعلم في حد ذاتها وتوليد تغذية راجعة مفيدة لكل المهتمين.
الوجاهة	مدى إمكانية المعلمين أو المقيمين الحكم مباشرة على قدرة الطالب على العمل في مهنة معينة دون الحاجة إلى الاستدلال على ذلك.
الشفافية	يجب أن تكون برامج تقييم الكفاءات واضحة ومفهومة لدى كل المهتمين.
العواقب التربوية	مدى تحقيق برامج تقييم الكفاءات آثار ايجابية في التعلم والتدريس وتقلل من الآثار السلبية.
إمكانية إعادة إنتاج القرارات	يجب أن تقوم القرارات حول الطلاب على عدة مقيمين وعدة مهام وعدة وضعيات.
إمكانية المقارنة	أن تكون مهام ومعايير وشروط العمل وإجراءات التقييم متسقة مع مميزاتها الأساسية.
التكاليف والفعالية	إمكانية تطوير وتنفيذ برامج تقييم الكفاءات بالنسبة لكل من الطلاب والمقيّمين.

المصدر: (Baartman, et al., 2006; Baartman, Prins, et al., 2007)

معايير الجودة المقدمة في هذا الإطار المفاهيمي ليست منفصلة عن بعضها البعض لأنها تعمل ضمن قالب واحد في برامج تقييم الكفاءات، كما أن كل أساليب التقييم الحديثة يجب أن تخضع لكل المعايير المحددة في هذا الإطار حتى تكون القرارات المتخذة حول الطلاب صائبة وموثوقة، فكلما توفرت هذه المعايير في برنامج التقييم كلما استطاع صانع القرار (مدرس، مسيّر...) التأكد من ملاءمته للاستدلال على كفاءات الطلاب، ولكن يمكن أن تختلف درجة توفر كل معيار حسب أساليب التقييم المدرجة ضمن برنامج الكفاءات.

خاتمة:

هدف هذا المقال إلى تقديم إطار مفاهيمي شامل ومتكامل لمعايير جودة برامج تقييم الكفاءات التي انبثقت من ثقافة التقييم، والتي بدورها أعادت النظر في ممارسات التقييم بعدما سيطرت ثقافة العملية الاختبارية مدة زمنية غير قصيرة، فقد أدت ثقافة التقييم إلى التوجه نحو إدماج التدريس والتقييم والتعلم لتحديد التقدّم الذي أحرزه الطالب لتأكيد أن التقييم عملية دينامية مستمرة ومندمجة لا تركز فقط على فترات معينة من عملية التعلم، ويكون فيها الطالب عنصرا فعالا في عملية التقييم من خلاله مشاركته في تطوير التقييم وإعداد المعايير وفي تقييم أعماله وأعمال أقرانه. وعلى أساس ثقافة التقييم طوّرت أساليب تقييم حديثة تتميّز فيها المهام بالواقعية والتعقيد، حيث تسمح باستثارة مهارات التفكير العليا، وتراعي في عملية التقييم كلا من النواتج النهائية والعمليات المؤدية إلى النواتج مع التركيز على الكفاءات متعددة الأبعاد كالمهارات المعرفية وما وراء المعرفية والوجدانية والحركية.

تتلاءم برامج تقييم الكفاءات مع ثقافة التقييم التي تعتبر كلا من التقييمات التقليدية والتقييمات الحديثة متكاملة الأدوار وغير متناقضة، كما لا تكتفي باستخدام أسلوب تقييم واحد بل مزيج من أساليب التقييم التقليدية والحديثة، فبرامج تقييم الكفاءات هي مزيج متكامل من أساليب التقييم التقليدية والحديثة. بالإضافة إلى عدم كفاية معياري الصدق والثبات الكلاسيكيين في تقييم جودة برامج تقييم الكفاءات، لذا تتطلب تكييفها حتى تتلاءم مع متطلبات أساليب التقييم الحديثة، وضرورة تقييم جودة برامج تقييم الكفاءات على أساس معايير جديدة وواسعة.

وفي إطار أدبيات التقييمات الحديثة تم تناول (12) معيارا لجودة برامج تقييم الكفاءات التي وردت في أعمال بارتمان وآخرون (Baartman, et al., 2006; Baartman, Prins, et al., 2007)، وفي ضوء هذا الإطار الشامل والمتكامل لمعايير الجودة التي استمدت في الأصل من ثقافة التقييم، تم وصف كل معيار جودة باختصار بمراعاة الدقة في التحديد حتى تُقدّم للمهتمين بتقييم جودة برامج تقييم الكفاءات أداة مفيدة تساعد على إعداد برامج تقييم ذات مستويات جودة عالية.

يمكن أن يكون هذا الاطار مفيدا للباحثين والمختصين في مجال تقييم جودة برامج التقييم من خلال إعداد أدوات تقرير ذاتية يتم على أساسها جمع معلومات عن برامج تقييم الكفاءات المستخدمة في الممارسات العملية، وباتباع هذه المعايير يمكن تقديم تغذية راجعة للمهتمين (مدرسين، مسيرين) داخل المدرسة لتحسين وتطوير وضمان جودة برامج التقويم المستخدمة، فالمعلومات التي تقدّمها أدوات تقييم برامج التقييم المعتمدة على معايير الجودة مفيدة للمهتمين أنفسهم ولصانعي القرارات لتصميم برامج تقييم ملائمة وذات جودة عالية، وتقديم مقترحات وتحسينات أكثر ابداعية وملاءمة.

قائمة المراجع:

. طباع، فاروق. (2017). انتقدات استخدام مصطلح الكفاءات في الممارسات التربوية المربطة بالتقويم. مجلة العلوم الاجتماعية، 24، 162-177.

. علام، صلاح الدين محمود. (2004). التقويم التربوي البديل: أسسه النظرية والمنهجية وتطبيقاته الميدانية. القاهرة: دار الفكر العربي.

. American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington: DC: Author.

- 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 (2), 153-177.

- Baartman, L. K., Bastieans, T. J., Kirschner, P. A., & Van der Vleuten, C. P. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review*, 2 (2), 114–129.

- Baartman, L. K., Gulikers, J., & Dijkstra, A. (2013). Factors influencing assessment quality in higher vocational education. *Assessment & Evaluation in Higher Education, 38* (8), 978-997.

- Baartman, L. K., Prins, F. J., Kirschner, P. A., & van der Vleuten, C. P. (2007). Determining the quality of competence assessment programs: A self-evaluation procedure. *Studies in Educational Evaluation*, 33 (3-4), 258-281.

- 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 (3), 206-216.

- Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum, & F. Dochy, *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3-30). Boston: Kluswer Academic.

- Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment/ In M. Segers, F. Dochy, & E. Cascallar, *Optimising new modes of assessment: In search of qualities and standards* (pp. 13-36). Boston: Kluwer Academic.

- Birenbaum, M., & Dochy, F. (1996). *Alternatives in assessment of achievement, learning processes and prior knowledge*. Boston: Kluwer Academic.

- Boud, D., & Falchikov, N. (2007). *Rethinking assessment in higher education: Learning for the longer term*. Oxon: Routledge.

- 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.

- Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educational and Psychological Measurement*, 57 (3), 373-399.

- De Ketele, J. M., & Gerard, F. M. (2005). La validation des épreuves selon l'approche par compétences. *Mesure et Évaluation en Éducation, 28* (3), 1-26.

- Dierick, S., & Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation*, 27 (4), 307–329.

- Dochy, F. (2001). A new assessment era: different needs, new chalenges. *Research Dialogue in Learning and Instruction,* 2, 11-20.

- 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.

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

- Gulikers, J. T., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research & Design*, 52 (3), 67-87.

- Gulikers, J., Biemans, H., & Mulder, M. (2009). Developer, teacher, student and employer evaluations of competence-based assessment quality. *Studies in Educational Evaluation*, 35 (2-3), 110-119.

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

- Jonsson, A., Baartman, L. K., & Lennung, S. A. (2009). Estimating the quality of performance assessments: The case of an 'interactive examination' for teacher competencies. *Learning Environment Resource*, 12 (3), 225-241.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20 (8), 15-21.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23 (2), 13-23.
- 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.
- Miller, M. D., & Linn, R. L. (2000). Validation of performance-based assessments. *Applied Psychological Measurement*, 24 (4), 367-378.
- Roegiers, X. (2000). *Pour une pédagogie de l'intégration*. Bruxelles: Be Boeck.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Bruxelles: De Boeck.
- Segers, M., Dochy, F., & Cascallar, E. (2003). *Optimizing new modes of assessment: In search for qualities and standards*. Boston: Kluwer Academic.
- 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.
- Tillema, H., Leenknecht, M., & Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning. *Studies in Educational Evaluation*, 37 (1), 25–34.
- Van der Vleuten, C. P. & Schuwirth, L. W. (2005). Assessing professional competence: From methods to programmes. *Medical Education*, 39 (3), 309-317.