

Analyse critique des démarches de vérification des qualités psychométriques du Questionnaire for Teacher Interaction (QTI)

Ibtissem Ben Alaya* – Benalayaibtissem@gmail.com

Vincent Grenon* – Vincent.Grenon@usherbrooke.ca

Naila Bali** – naila_bali@yahoo.fr

Jean François Desbiens* – Jean-Francois.Desbiens@usherbrooke.ca

*Université de Sherbrooke, Faculté d'éducation

Centre de recherche interuniversitaire sur la formation et la profession enseignante (CRIFPE)

**Institut Supérieur du Sport et de l'Éducation Physique de Ksar Saïd, Université de la Manouba.
Laboratoire Optimisation de la Performance Sportive

Pour citer cet article : Ben Alaya, I., Grenon, V., Bali, N., & Desbiens, J.F. (2018). Analyse critique des démarches de vérification des qualités psychométriques du Questionnaire for Teacher Interaction (QTI). *Évaluer. Journal international de recherche en éducation et formation*, 4(2), 3-26.

Résumé

Nombre d'études visant à vérifier les qualités psychométriques des questionnaires en éducation ne semblent pas faire un usage adéquat des analyses statistiques ou aboutir à une interprétation appropriée des résultats trouvés (Beavers *et al.*, 2013). Des travaux réalisés avec le Questionnaire on Teacher Interaction (QTI) n'échappent pas à ce constat bien que cet outil ait fait l'objet de plus de 1500 publications scientifiques au cours de ces deux dernières décennies. Cet article s'inscrit dans ce cadre et a pour objet de décrire et d'analyser de manière critique, différents protocoles de vérification des qualités psychométriques utilisés avec ce questionnaire. Pour ce faire, un corpus de vingt-trois études, publiées entre 1985 et 2016, a été constitué selon une démarche systématique de recherche documentaire. L'analyse des données a été menée à l'aide d'une grille d'analyse permettant d'identifier, pour chaque étude, le contexte, le type de validité et de fidélité concerné, les modèles statistiques, les caractéristiques de l'échantillon et la version du QTI faisant l'objet de la validation. Ces éléments ont mené à deux comparaisons systématiques, une première entre les études ciblées et une seconde visant à établir leur adéquation par rapport aux principes généraux recommandés au plan méthodologique. Nous avons formulé, pour chaque lacune identifiée, quelques propositions dans le but d'améliorer et de systématiser les futures démarches de validation d'un questionnaire en sciences humaines, et en particulier le QTI.

Mots clés

QTI ; Qualités psychométriques ; Validité ; Fidélité ; Critique.

Abstract

Many studies that are conducted to test psychometric properties of questionnaires used in educational settings do not seem to use correctly the statistical analyzes or to arrive to an appropriate interpretation of the results achieved (Beavers *et al.* 2013). Research carried out with the Questionnaire on Teacher Interaction (QTI) does not escape this observation although this tool has been the subject of more than 1000 peer-reviewed publications over the last two decades. The purpose of this study is to describe and analyze in a critical way different protocols of verification of the psychometric qualities that were used with this questionnaire. To do this, a corpus of twenty-three studies published between 1985 and 2016 was constituted according to a systematic process of documentary research. The analysis of the data was carried out using framework created to identify, for each study, the context, the type of validity and fidelity concerned, the types of statistical techniques, the characteristics of the samples and the version of the QTI being validated. These elements led to two systematic comparisons, one between the targeted studies and the other to establish their adequacy with the general principles recommended on methodology. We have formulated, for each identified deficiency, some proposals to improve and systematize future validation processes of the QTI.

Keywords

QTI; Psychometric properties; Validity; Fidelity; Critical

1. Introduction et problématique

Les questionnaires psychométriques font partie des outils de collecte de données les plus utilisés en sciences humaines (Huot, 2003). Or, depuis l'avènement des logiciels statistiques, les questionnaires psychométriques ont été fortement critiqués par plusieurs équipes de recherche (Beavers *et al.*, 2013; Bourque *et al.*, 2006 ; Fabrigar, Visser & Browne, 1997; Ferguson & Cox, 1993; Field, 2013 ; Osborne & Costello, 2009). Le retour sur des études récentes en psychologie (Fabrigar *et al.*, 1997; Russell, 2002), en recherche organisationnelle (Conway & Huffcutt, 2003), en communication (Park, Dailey & Lemus, 2002), en éducation (Henson, Capraro & Capraro, 2001; Pohlmann, 2004), etc. montre que la plupart des démarches de validité et de fidélité sont entachées plus ou moins sévèrement par l'usage inadéquat des analyses factorielles, l'interprétation non appropriée des résultats trouvés ou le choix difficilement défendable de certaines méthodes d'analyse de données (Beavers *et al.*, 2013; Bourque *et al.*, 2006). Ces études peuvent remettre en question la crédibilité de certains questionnaires et des résultats qui en découlent. Toutefois, ceci ne signifie pas que tous les questionnaires ne présentent pas des qualités psychométriques suffisantes. Sinon, on devrait arrêter d'utiliser les questionnaires et choisir d'autres moyens pour faire avancer la connaissance. Il est par ailleurs important de vérifier les qualités psychométriques de chaque questionnaire avant de l'utiliser.

Nous avons traité, dans cet article, le cas de l'un des questionnaires les plus utilisés en éducation pour étudier l'environnement d'apprentissage et sa dimension relationnelle (den Brok *et al.*, 2002 ; Dumas, 2009), à savoir le *Questionnaire on Teacher Interaction* (QTI). Depuis sa première diffusion en 1985 (Wubbels, Creton & Houymayers, 1985), le QTI a été mobilisé au sein de plus de 1500 publications scientifiques répertoriées dans Google Scholar 2016 (Google Scholar, 17-05-2017) et traduit dans plus de 20 langues. Il a fait aussi l'objet de plusieurs démarches de validation dans plus de 40 pays (den Brok *et al.*, 2003a), à différents niveaux d'études (primaire, secondaire et universitaire) et dans différentes disciplines scolaires (ex. : mathématiques, sciences, anglais, etc.) (Sivan & Chan, 2013). Ces démarches de validation expliquent vraisemblablement l'existence de différentes versions du QTI depuis la version hollandaise originale de 77 items (Wubbels, 1985) jusqu'à la version canadienne française la plus courte composée de 32 items (Lapointe & Legault, 1999). Peu importe la version utilisée, ce questionnaire est censé permettre de déterminer comment les élèves du secondaire perçoivent les comportements interpersonnels de leurs enseignants et de regrouper leurs perceptions selon deux dimensions ou construits, à savoir le soutien et le contrôle. Ces deux dimensions sont reconnues et acceptées comme marqueurs universels de l'interaction enseignant-enseigné (Fisher & Rickards, 1998 ; Dumas, 2009). En effet, la structure en circomplexe héritée des travaux de Leary (1957) et les fondements de la communication interpersonnelle sur lesquels a été basé le QTI offrent une lecture simplifiée de la relation enseignant-enseigné qui a facilité son utilisation, aussi bien par des chercheurs et des enseignants que par des pédagogues et des responsables du secteur éducatif (Cohen, McCabe, Michelli & Pickeral, 2009 ; Debarbieux, 2015).

Néanmoins, en dépit de la vaste diffusion du QTI dans le monde, l'utilisation soutenue de cet instrument doit être questionnée sur la base d'apparentes failles méthodologiques susceptibles de remettre en question ses qualités psychométriques. Entre autres, l'utilisation inadéquate de

certaines analyses statistiques de ce questionnaire peut conduire à des appréciations biaisées de sa fidélité et de sa validité, ce qui risque d'affaiblir plausiblement la robustesse des résultats produits depuis plus de 25 ans en utilisant ce questionnaire.

Dans les études antérieures, nous n'avons trouvé qu'une seule analyse rétrospective des démarches de vérification de ses qualités psychométriques. Cette analyse relativement superficielle a été publiée par den Brok *et al.* (2003a). En comparant des études publiées dans six pays différents, ces auteurs ont relevé plusieurs faiblesses susceptibles de nuire à la validité et à la fidélité de l'instrument (ex. choix non approprié des méthodes d'extraction des facteurs, interprétation non justifiée des résultats trouvés, choix non justifié des méthodes d'analyses statistiques, etc.). Toutefois, cette étude n'avait pas pour objectif de critiquer le QTI. Les faiblesses de cet instrument ont été présentées d'une façon générale, à titre informatif et sans avoir recours à des exemples ciblés ni à une démarche de critique systématique. Par ailleurs, les chercheurs continuent à utiliser ce questionnaire en prenant pour acquis qu'il possède des qualités psychométriques jugées suffisantes et cela depuis plus de 25 ans (Dumas, 2009). Les programmes de formation continuent aussi à accepter la perspective bidimensionnelle de la relation enseignant-élève partagée par les études basées sur le QTI.

Ainsi, nous proposons de produire une analyse critique approfondie des démarches utilisées dans la documentation scientifique pour étudier la validité et la fidélité du QTI. De manière plus précise, l'attention sera essentiellement dirigée vers la validité de construit (factorielle et structurale) et la cohérence interne et temporelle comme estimation de la fidélité. Le but de cet article est de vérifier l'écart entre ce qui est recommandé par les experts des analyses factorielles (ex. Beavers *et al.*, 2013; Fabrigar *et al.*, 1997; Ferguson & Cox, 1993; Field, 2013) et ce qui est présenté concrètement dans les démarches de vérification de la validité et de la fidélité du QTI.

2. Méthodologie

Dans cette section, il est d'abord question de décrire la démarche en cinq étapes inspirée de Paillé (2007) qui nous a permis de sélectionner notre corpus d'étude. Par la suite, nous exposons d'une façon détaillée comment nous avons collecté les données. Dans la première étape, nous avons collecté, pour l'ensemble des articles analysés, les informations décrivant le contexte d'étude (la version, l'objectif de l'étude). Nous avons, par la suite, vérifié si les prémisses relatives à l'échantillon et à la matrice des corrélations ont été respectées. L'attention a été, dans une troisième étape, accordée à la démarche de validité factorielle, en commençant par la méthode et le critère d'extraction jusqu'à traiter de la complexité des variables. Nous avons vérifié par la suite si, pour chaque étude, la validité structurale a été menée selon ce qui est recommandé par Leary (1957). Nous avons clôturé cette partie par la collecte des données relatives au type de coefficient de fidélité utilisé, sans oublier le seuil accepté et la façon dont les résultats ont été interprétés.

2.1. La sélection du corpus étudié

Notre corpus compte 23 études scientifiques rédigées en anglais ou en français et publiées entre 1985 et 2016 dans des revues scientifiques avec comité de lecture. Ce corpus correspond

donc à notre échantillon qui a été élaboré en plusieurs phases et selon des critères bien déterminés.

- 1) Dans la première étape, il a été question de circonscrire notre recherche à l'aide de sept concepts clés, à savoir : QTI ; validité ; fidélité ; qualité psychométrique ; analyse factorielle ; structurale et alpha de Cronbach.
- 2) Par la suite, une grille d'analyse a été utilisée afin d'établir plusieurs équations de recherche ¹ à partir de nos concepts clés ainsi que leur traduction en langue anglaise. Pour augmenter les chances d'avoir des publications proches de notre sujet d'étude, plusieurs moteurs de recherches, qu'ils soient spécifiques au domaine de l'éducation ou à caractère plus général, ont été consultés (ex. Google Scholar, Google Books, Eric, Cairn-Info, Education Source, etc.).
- 3) Les différentes équations de recherche ont permis d'avoir une liste conséquente de publications de différentes formes (articles de revues, thèses, forums, livres, actes de colloques, etc.), publiés dans différentes langues (français, anglais, néerlandais, etc.) et différentes périodes et qui sont plus au moins proches de ce qui était recherché.
- 4) À partir de cette liste, nous avons effectué un tri selon des critères d'inclusion et d'exclusion bien établis. Le premier critère concerne **la période de publication** qui s'étale, tel qu'exprimé plus haut, de la date de l'adaptation du QTI en contexte scolaire (Wubbels, 1985) jusqu'en 2016. La **forme de la publication** correspond à notre deuxième critère. Précisément, seuls les articles publiés dans des revues avec des comités de lecture ont été conservés. Les livres, par exemple, ont été éliminés puisqu'ils ne sont pas systématiquement révisés par des pairs. **L'accessibilité linguistique** a formé notre troisième critère, vu que plusieurs études ont été publiées en grec, en turc, en chinois, en suédois, en russe, en néerlandais, etc. Seuls les textes en français ou en anglais ont été retenus dans notre corpus. Néanmoins, dans le but de former le corpus le plus représentatif possible de la documentation existante, pour chaque version linguistique accessible, nous avons sélectionné au **moins deux pays différents** (ex. des études en Australie, aux E.U et à Singapour pour la version anglaise, la France et le Québec ont formé des exemples d'études pour la version française). Nous avons également essayé de constituer un corpus qui renferme des études menées sur **chaque continent** à travers le monde comme Wubbels et Levy (1993) pour l'Amérique, Passini *et al.* (2015) pour l'Europe, Kim, Fisher et Fraser (2000) pour l'Asie et l'Océanie, etc. L'ordre **scolaire** est un autre critère mis à profit pour

¹ L'équation de recherche est une méthode qui permet d'obtenir plus de résultats sur internet selon des critères d'inclusion et d'exclusion bien déterminés. Avant de procéder par cette méthode, il est important de cerner les concepts clés du sujet de recherche. Dans notre article nous en avons défini 7. En liant ces concepts par les opérateurs booléens (et (+), exclu (-)) nous définissons plusieurs équations de recherche susceptibles de nous faire ressortir des publications en lien avec notre sujet. Les publications trouvées doivent être triées par la suite pour conserver seulement celles qui respectent nos critères d'exclusion et d'inclusion.

NB. Au lieu de mettre validité, validation, valide, on peut mettre valid* pour avoir les trois mots.

Exemples d'équations sur Éric : QTI+ valid* + fidél* ; QTI+ analyse factorielle + valid* ; QTI + fidél* + alpha de Cronbach ; QTI + valid* + reabil*+ Factorial ; etc.

Refaire les mêmes équations avec d'autres moteurs de recherche comme Ebsco, CAIRN, etc.

Chaque fois qu'on change l'équation de recherche, les résultats changent (nombres d'articles, de nouveaux titres apparaissent à chaque fois).

former le corpus le plus représentatif possible de la réalité. En effet, vu que le QTI a été administré auprès de différentes catégories d'âge, nous avons essayé de choisir au moins deux publications réalisées pour chaque niveau scolaire (primaire, secondaire et universitaire). Le QTI a été validé dans plusieurs **disciplines scolaires**. Ceci explique que notre corpus renferme aussi des démarches de validation dans des classes de mathématiques, de sciences, d'éducation physique et sportive, etc.

- 5) Très tôt, il est apparu que **les qualités psychométriques** du QTI ont été vérifiées différemment d'une étude à une autre. Toutefois, en raison des limites d'espace, nous avons choisi d'analyser seulement les démarches traitant de **la validité de construit** (factorielle et structurale), de **la cohérence interne** (alpha de Cronbach ou coefficient de corrélation intra-classe) et de la **fidélité temporelle**. Cela a conditionné le choix des articles à analyser.

La démarche décrite ci-dessus nous a permis de sélectionner un corpus de 23 articles scientifiques (annexe 1) qui respectent tous nos critères d'inclusion et qui forment un échantillon diversifié en termes de niveau d'études, langue et période de publication, contexte de validation et localisation géographique.

2.2. La collecte et l'analyse des données

Nous avons préparé une grille d'analyse pour collecter les données pertinentes de chaque étude. Les lignes de notre grille correspondent aux dimensions qu'on souhaite analyser. Le choix de ces dimensions a été basé sur les guides méthodologiques traitant les principes de base de la validité factorielle, structurale et de la fidélité (Beavers *et al.*, 2013; Fabrigar *et al.*, 1997; Ferguson & Cox, 1993; Field, 2013 ; Osborne & Costello, 2009; Reise, Waller & Comrey, 2000).

La première partie de notre grille renferme des informations générales sur le QTI et le contexte de l'étude telles que la version utilisée, le pays, le niveau scolaire, l'objectif de l'étude, la discipline scolaire, etc. La prise en compte de ces informations nous a permis de sélectionner un échantillon des études réalisées sur la base d'un éventail plus riche des contextes dans lesquels elles ont été produites, tout en ayant une vue d'ensemble des ajustements apportés par les chercheurs d'une version à une autre du QTI.

La deuxième partie de cette grille a été consacrée aux prémisses de la validité factorielle, structurale et de la fidélité, en particulier de la consistance interne. En nous référant toujours aux principes méthodologiques recommandés, une attention particulière a été portée à l'échantillon (ex. : technique d'échantillonnage, taille) et à la matrice de corrélation comme étant deux prémisses essentielles dans n'importe quelle démarche de validation factorielle (Beavers *et al.*, 2013; Field, 2009). En ce qui concerne l'échantillon, notons que « *an inadequate sample size can be detrimental to the factor analytic process and produce unreliable, and therefore, non-valid*

results» (Beavers *et al.*, 2013). Il est par ailleurs recommandé d'utiliser les ratios² et de vérifier les communalités (qualité de représentation ou chargement factoriel) pour justifier la taille de l'échantillon. En général, un échantillon de 300 sujets produit une solution factorielle stable (Field, 2013, p.675). La technique de l'échantillonnage et les caractéristiques de l'échantillon doivent être bien décrites et justifiées. La matrice de corrélation doit, quant à elle, indiquer qu'elle peut se prêter à des analyses factorielles (Durand, 2005). Les corrélations trop élevées ($r > .80$) doivent être supprimées pour éviter le problème de multi-colinéarité et de singularité (Field, 2013, p. 685). Un déterminant de la matrice positif et différent de zéro indique qu'une solution factorielle est mathématiquement possible (Pett, Lackey & Sullivan, 2003). Selon Field (Field, 2013), le déterminant doit avoir un seuil minimal de 10^{-5} . Les corrélations trop faibles présentent quant à elles le risque d'avoir une matrice identité. Le test de Bartlett peut être utilisé pour vérifier ce risque. Si le test est significatif, on peut effectuer des analyses factorielles (Beavers *et al.*, 2013; Field, 2013, p. 685). Idéalement, ces tests doivent être accompagnés du test de *Kaiser-Meyer-Olkin* (KMO) pour vérifier l'adéquation de la solution factorielle attendue (Beavers *et al.*, 2013).

La troisième ligne de notre grille concerne la validité factorielle. Puisque cette dernière est tributaire de la solution factorielle trouvée, nous avons collecté des données relatives à la méthode d'extraction utilisée, la justification du nombre de facteurs conservé, la rotation des facteurs, la suppression des variables complexes et l'interprétation des résultats trouvés.

En ce qui concerne les méthodes d'extraction, le choix d'une analyse factorielle exploratoire (AFE), confirmatoire (AFC) ou d'une analyse en composantes principales (ACP) dépend de l'objectif de l'étude (validation, revalidation), du contexte de l'étude (un nouveau contexte, un contexte qui a été déjà étudié), et des communalités (faibles ou élevées) (Bourque *et al.*, 2006; Fabrigar *et al.*, 1997; Ferguson & Cox, 1993). La documentation scientifique offre un large éventail de critères d'extraction permettant de prendre une décision quant au nombre de facteurs à extraire (ex. critère de Kaiser, 1960; Cattell, 1966 ; Horn, 1965) (Beavers *et al.*, 2013 ; Osborne & Costello, 2009; Field, 2013, p. 683). Précisons que le critère de Horn est classé de nos jours parmi les critères les plus rigoureux (Fabrigar *et al.*, 1997). Une fois le nombre de facteurs décidé, il est recommandé d'effectuer une rotation des axes en faisant attention aux variables complexes pour aboutir à la solution factorielle la plus appropriée. Le seuil de détermination d'une variable complexe est souvent fixé à une différence de chargement $< .2$ ($< .3$ selon des normes plus rigoureuses). Chaque variable complexe doit être retirée de l'analyse (Beavers *et al.*, 2013).

² Plusieurs chercheurs ont défini des ratios pour justifier la taille de l'échantillon adoptée. Kline (1986) et Nunnally (1978) par exemple, proposent le calcul du ratio N/P (N : taille de l'échantillon; P : le nombre des variables). Selon ces deux auteurs, la taille est dite suffisante si le ratio est situé entre 2.1 et 10.1. Vous pouvez consulter l'étude de Ferguson et Cox (1993) pour plus de détails sur ce sujet. D'autres chercheurs recommandent le recours aux communalités pour confirmer que la taille de l'échantillon est suffisante (ex. Selon Field (2013) : si les facteurs renferment 10 ou 12 variables avec des communalités de 0.4 et plus, un échantillon d'au moins 150 sujets est nécessaire. Pour des facteurs composés de peu de variables (comme le QTI : 8 facteurs de 4 variables chaque) ayant des communalités faibles à modérées, il est important d'avoir un échantillon d'au moins 300 sujets.

La validité structurale occupe la quatrième position dans notre grille. Dans le cas du QTI, cette forme de validité sous-entend la vérification de l'hypothèse de Leary (1957) qui suppose que, pour chaque échelle du QTI, les corrélations positives les plus fortes sont obtenues avec les échelles adjacentes et la corrélation négative la plus élevée avec l'échelle directement opposée dans le circomplexe. Partant de là, il est important de vérifier, pour chaque étude analysée, le type de coefficient de corrélation utilisé, le nombre d'hypothèses vraies, etc.

Notre grille a aussi renfermé une case qui concerne la fidélité de l'instrument en mettant l'accent sur le type de fidélité et de coefficient choisi, le seuil accepté et la justification du choix établi.

2.3. Le « Questionnaire on Teacher Interaction » (QTI)

Cet instrument de mesure s'appuie sur le modèle théorique en circomplexe de Leary (1957). Il prend aussi en considération les principes de la communication interpersonnelle de Watzlawick, Bavelas, Jackson et O'Hanlon (2011). Wubbels *et al.* (1985) ont créé la première version du QTI au milieu des années 1980 et ils l'ont adaptée au contexte scolaire pour permettre aux chercheurs d'étudier les comportements interpersonnels de l'enseignant en classe, tout en restant fidèles à la théorie communicationnelle et au modèle de Leary.

Cette première version contenait 77 items. Elle a été validée et administrée aux Pays-Bas en langue néerlandaise à partir d'un échantillon de 2407 élèves du secondaire (Wubbels et Leavy, 1993). Elle a démontré des qualités psychométriques jugées suffisantes grâce à des seuils d'alpha de Cronbach situés entre .77 et .87 et des coefficients de corrélation intra-classe situés entre .94 et .97 (Wubbels, 1985).

Les chercheurs ont tenté de réduire le nombre d'items dans le but de l'alléger et d'en faciliter l'administration. Par la suite, une version américaine de 64 items a été validée aux États-Unis (Wubbels et Levy, 1991). Quelques années plus tard, des chercheurs australiens (Fisher, Fraser, Wubbels et Brekelmans, 1993) ont réussi à réduire le QTI à 48 items. Cette version a été utilisée dans plusieurs contextes et disciplines scolaires pour arriver en 1999 à la version la plus courte de 32 items. Il s'agit d'une version en français validée par Lapointe et Legault (1999) au Québec. Le tableau 1 inspiré de l'étude de Passini *et al.* (2015) présente quelques informations à propos de certaines versions du QTI. En ce qui concerne la fidélité temporelle, les pionniers du QTI (Wubbels et Levy, 1989, 1993) ont démontré, en se basant sur le coefficient test-retest, qu'il n'est pas nécessaire d'administrer le QTI plus d'une fois par an pour avoir des résultats stables et statistiquement significatifs, pour autant que son administration soit faite quelques semaines après le début de l'année scolaire. Ce délai permet aux perceptions d'acquérir une certaine stabilité (Dumas, 2009). La validité structurale sous-entend la vérification de l'hypothèse de Leary (1957) présentée précédemment.

Tableau 1. Caractéristiques de quelques versions du QTI

Langue	Version	Taille de l'échantillon	Références	Alpha de Cronbach	Validité structurale
Néerlandais	77 items	1105	Wubbels <i>et al.</i> (1985)	> .74	Non vérifiée
Anglais (EU)	64 items	1606	Wubbels et Levy (1993)	> .76	Non vérifiée
Anglais (Australie)	48 items	3215	Rickards et Fisher (1999)	> .62	Non vérifiée
Français	64 items	799	Lapointe et Legault (1999)	> .68	Partiellement confirmée
Français	32 items	1764	Lapointe et Legault (1999)	> .62	Confirmée
Coréen	48 items	543	Kim <i>et al.</i> (2000)	> .61	Confirmée
Grec	64-Item	1973	Kyriakides (2005)	> .84	Confirmée

3. Résultats

Les 23 articles analysés présentent des points de convergence et de différence avec ce qui est recommandé par les experts en méthodologie et en statistiques en ce qui a trait à la taille et à la technique d'échantillonnage, la matrice de corrélation, la validité factorielle, la validité structurale et, enfin, la fidélité. Ces éléments sont traités en détail dans ce qui suit.

3.1. Constats généraux

Un nombre considérable d'articles analysés sont pauvres en informations relatives à la version du QTI utilisée ou à l'objectif de la recherche (première ligne de la grille). Dans au moins quatre articles parmi les 23 analysés, aucune information n'a été fournie sur la version du QTI utilisée (Henderson, Fisher & Fraser, 2000; Koul & Fisher, 2006, 2005; Levy *et al.*, 2003). Henderson *et al.* (2000) et Wubbels (1985) n'identifient pas, quant à eux, l'objectif de leurs études. Ils ne précisent pas s'il s'agit de la validation d'une nouvelle version du QTI ou d'une revalidation d'une version existante, etc. Or, ces informations forment des arguments importants pour justifier le choix de la méthode d'extraction, de la taille de l'échantillon, etc. À défaut de les connaître, il sera difficile de juger la qualité de la démarche d'analyse effectuée.

3.2. Vérification des prémisses

3.2.1. Taille et technique d'échantillonnage

Sur les 23 articles analysés, seulement deux (2 / 23) (den Brok, Levy, Wubbels & Rodriguez, 2003; Wei, den Brok & Zhou, 2009) n'ont pas respecté la taille minimale d'échantillon recommandée pour amorcer des analyses factorielles. Ces recherches comptent respectivement 78 et 160 sujets. Néanmoins, la taille de l'échantillon utilisée, n'a été justifiée que par Wei *et al.* (2009) (1 / 23). Ces derniers sont les seuls qui ont justifié la taille de leur échantillon en se référant aux ratios de taille et de sexe. Aucune information n'a été fournie à ce sujet dans les 22 autres articles analysés.

En ce qui concerne les techniques d'échantillonnage, quatre études ont été basées sur des échantillons par convenance (Sivan, Chan & Kwan, 2014; Sivan & Chan, Telli, den Brok & Cakiroglu, 2007; Wei *et al.*, 2009). den Brok *et al.*(2003) et Wubbels & Levy (1980) indiquent avoir recouru à des échantillons non aléatoires, sans toutefois fournir plus détails. Pour le reste des articles, aucune information n'a été fournie quant aux techniques d'échantillonnage adoptées. Seulement cinq articles présentent une description de l'échantillon en termes des caractéristiques des participants (sexe, âge, discipline scolaire et niveau d'études) (den Brok *et al.*, 2003; Fisher, *et al.*, 2003; Goh & Fraser; Telli *et al.*, 2007; Wei *et al.*, 2009; Wubbels, 1985).

3.2.2. Matrice de corrélations

La qualité de la matrice de corrélations est un des éléments à surveiller lorsqu'on fait des analyses factorielles. La vérification du déterminant de la matrice, du test de Bartlett et de l'indice de KMO est toujours recommandée pour s'assurer que la matrice se prête bien à des analyses factorielles (Bourque *et al.*, 2006).

Sur les 23 articles analysés, seulement deux (Scott & Fisher, 2004; Sivan, Chan et Kwan, 2014) fournissent des informations à ce sujet. Pour les autres, on ne sait pas si la matrice utilisée est une matrice de corrélations ou de covariances. Scott et Fisher (2004) ont rapporté avoir supprimé deux items de leur matrice de corrélations qui présentaient des corrélations très faibles et ont, par la suite, testé la nouvelle matrice sans ces deux items. Cependant, ces auteurs n'ont pas rapporté s'il s'agit d'une matrice identité (test de Bartlett et déterminant) et si l'indice du KMO était suffisant. La seule étude qui a traité de ces différents aspects concernant la matrice est celle de Sivan, Chan et Kwan (2014).

3.3. Validité factorielle

3.3.1. Méthodes d'extraction

Les analyses factorielles exploratoires peuvent être effectuées à partir d'un modèle en composantes principales ou d'un modèle en facteurs communs selon l'objectif de l'étude, le contexte étudié, les communalités, etc. (Beavers *et al.*, 2013; Osborne & Costello, 2009). Ce choix doit toujours être justifié.

Nous avons pu identifier les méthodes d'extraction utilisées dans seulement six articles parmi les 23 analysés. Kokkinos, Charalambous & Davazoglou (2009) ont effectué des ACP suivies d'AFC. Sans justifier leur choix et bien qu'il s'agisse d'une première validation du QTI en Chine, Wei *et al.* (2009) ont choisi la méthode confirmatoire (AFC). Ces auteurs se sont basés sur l'hypothèse du modèle deux dimensions indépendantes et ont vérifié la corrélation entre ces deux dimensions avec le logiciel Mplus. Telli, den Brok & Cakiroglu (2007) ont opté pour des AFE afin de vérifier la validité d'une version adaptée en turc du QTI. Des ACP ont été priorisées par Fraser, Aldridge & Soerjaningsih (2010) et Sivan et Chan (2013) pour justifier une nouvelle adaptation du QTI, respectivement au contexte universitaire et au contexte du primaire. Trois autres articles ont mentionné simplement le recours à des analyses factorielles, sans préciser la méthode d'extraction choisie (Levy, Wubbels, den Brok & Brekelmans, 2003; Wubbels, 1985; Wubbels & Levy, 1989).

3.3.2. Critère d'extraction

Pour justifier la validité factorielle, il est important de préciser le nombre de facteurs à extraire en se basant sur un ou plusieurs critères d'extraction (Field, 2009).

Seulement dans quatre articles parmi les 23 analysés, les auteurs se sont basés sur au moins un critère d'extraction pour justifier le nombre de facteurs à extraire. Par conséquent, dans les articles restants (19 / 23), peu importe la nature de leur étude (nouveau contexte ou contexte étudié et exploré dans le passé), les chercheurs supposent l'existence de deux facteurs (soutien et contrôle) comme le postule le modèle théorique sur lequel a été basé le QTI, sans toutefois le vérifier méthodiquement.

Pour les quatre articles ayant traité ce point, on remarque que la décision est souvent étonnante. Par exemple, Sivan et Chan (2013) ont trouvé dix facteurs avec la méthode de Kaiser et trois avec celle de Horn. Cependant, et contrairement à ce qui est attendu, ils ont effectué une extraction forcée de huit facteurs, vu que le modèle théorique du QTI contient huit échelles.

Pour leur part, Kokkinos, Charalambous & Davazoglou (2009) se sont référés aux critères de Kaiser (1960) et de Cattell (1966) qui ont donné respectivement treize et trois facteurs. Contrairement à ce qui est attendu, ces auteurs ont extrait sept facteurs. Cette décision a été expliquée par le fait que sept parmi les huit échelles du QTI ont de bonnes consistances internes. Autrement dit, ces auteurs ont aussi tenu pour acquis que la solution factorielle comprend huit échelles conformément au modèle théorique sous-jacent, et ils ont extrait les sept facteurs ayant de bonnes consistances internes. Toutefois, on remarque que l'alpha de Cronbach est inférieur à .70 pour au moins cinq facteurs. Levy *et al.* (2003) ont respecté le critère de Kaiser qui a donné deux facteurs. Wubbels et Levy (1989) ont aussi justifié le choix de deux facteurs en se basant sur les critères de Kaiser et Cattell qui ont donné le même résultat, soit deux facteurs.

3.3.3. Variables complexes et rotation

Rappelons que la solution factorielle résultante d'une AFE nécessite la plupart du temps une rotation des axes afin de faciliter l'interprétation des résultats et la suppression d'items complexes qui peuvent causer problème lors de la définition des facteurs latents. Seulement quatre articles sur 23 ont précisé cette information. Sivan et Chan (2013) ont utilisé la rotation oblique. On ne sait pas de quel type de rotation oblique il s'agit. Fraser *et al.* (2010) et Levy *et al.* (2003) ont utilisé une rotation orthogonale de type Varimax. Wubbels et Levy (1989) ont précisé, quant à eux, qu'ils ont effectué une rotation à la main sans l'identifier. Dans les quatre cas présentés, aucune justification n'a été fournie quant au choix du type de rotation.

La plupart des auteurs tiennent pour acquis l'existence de deux facteurs et de huit échelles dans leur démarche de validation. Ainsi, ils ignorent volontairement ou involontairement la complexité des items pour conserver les huit échelles. D'ailleurs, sur les 23 études analysées, seuls Fraser *et al.* (2010), Kokkinos *et al.* (2009) et Sivan *et al.* (2014) ont vérifié cette condition. Pour les autres études, la plupart du temps, les tableaux de la solution factorielle ne sont pas présentés, ce qui ne permet pas d'apprécier les solutions produites ni d'identifier et d'interpréter les situations où il y a effectivement des items complexes. Goh et Fraser (1996),

Wei *et al.* (2009) ainsi que Wubbels et Levy (1989) font partie des rares équipes qui fournissent ces tableaux sans toutefois prendre en considération la notion de complexité. Si on analyse leurs solutions en fixant 0,20 comme seuil de complexité, au moins une échelle doit être enlevée dans chaque solution produite, ce qui remet en question la validation des huit échelles du QTI.

3.4. Validité structurale

Rappelons que la validité structurale est basée sur l'étude des corrélations entre les huit échelles du QTI (Lapointe & Legault, 1999). Sur les 23 articles analysés, huit n'ont pas vérifié ce type de validité (Fisher, Fraser et Rickards., 1997 ; Fraser et al., 2010; Henderson et al., 2000; Koul & Fisher, 2005; Maulana et al., 2012; Rickards & Fisher, 2010; Telli et al., 2017; Wei et al., 2009). Sur les 15 ayant vérifié ce type de validité, plusieurs publications ne fournissent pas les tableaux de corrélations, ce qui ne permet pas de porter un jugement sur les résultats trouvés (Goh & Fraser, 1996 ; Koul & Fisher, 2006; Wubbels & Levy, 1989). D'autres comme Coll, Taylor & Fisher (2002), Kokkinos *et al.* (2009) et Sivan *et al.* (2014) fournissent des tableaux incomplets, présentant seulement l'intersection entre sept échelles du QTI, en enlevant la huitième, celle de l'indulgence. L'absence de cette échelle n'est pas justifiée. Sivan & Chan (2013), quant à eux, confirment la structure en circomplexe, bien que toute une partie de leur tableau ne contienne aucun signe négatif, alors que le modèle en circomplexe suppose l'existence d'échelles opposées.

Nous avons remarqué que les tableaux de corrélations ont été présentés dans environ la moitié des cas. Toutefois, dans ces différents articles, seul celui de Legault et Lapointe (1999) confirme la validité structurale en calculant le nombre d'hypothèses vraies, soit 17 sur 24 selon les auteurs. Tous les autres articles ont confirmé la validité structurale en donnant un exemple d'hypothèse vraie et en invitant le lecteur à aller voir le tableau. En calculant le nombre d'hypothèses vraies pour chacun de ces articles, on trouve que pour la plupart, il s'agit de 16 hypothèses vraies sur 24 (Lee, Fraser & Fisher, 2003; Scott & Fisher, 2004 ; Sivan *et al.*, 2014).

Le tableau de Kim *et al.* (2000) présente seulement 12 hypothèses vraies sur 24, tandis que 22 hypothèses ont été confirmées pour les études de Passini *et al.* (2015) et de Wubbels et Levy (1989). Bien que le modèle en circomplexe suppose la confirmation de l'hypothèse de Leary pour toutes les échelles, les auteurs choisissent d'accepter la validité structurale en dépit de la présence de plusieurs irrégularités et/ou hypothèses non confirmées (ex. Coll *et al.*, 2002 ; Levy *et al.*, 2003 ; Scott & Fisher, 2004). Aucun critère n'est fourni quant au pourcentage ou au nombre d'hypothèses non confirmées à partir duquel la validité structurale doit être rejetée. Par conséquent, la décision rapportée n'est jamais justifiée. Elle est, la plupart du temps, justifiée dès que quelques exemples tiennent la route et tendent à confirmer l'hypothèse de Leary.

3.5. Fidélité

La fidélité d'un instrument de mesure, d'un test ou d'une échelle (en anglais : « *reliability* ») fait référence au degré de sa précision (Nunnally, 1978), de sa constance (Fortin, 1994) et de sa corrélation avec elle-même (Demeuse et Henry, 2004). Bien qu'il existe différentes manières d'apprécier cette qualité psychométrique, en pratique, nous constatons que le type de mesure de fidélité employé n'est pas identifié dans neuf articles (9 /23) (Coll *et al.*, 2002; den Brok *et al.*,

2003 ; Fisher *et al.*, 1997 ; Kim *et al.*, 2010 ; Koul & Fisher, 2005; Levy *et al.*, 2003; Rickards & Fisher, 2000; Telli *et al.*, 2007; Wubbels *et al.*, 1985). Il ressort que quatre équipes ont même considéré l'alpha de Cronbach comme un indice de validité du QTI à savoir : Fisher *et al.* (1997) ; Henderson *et al.* (2000) ; Koul & Fisher (2005) ; Maulana *et al.* (2012). Entre autres, ces chercheurs se sont basés sur l'alpha de Cronbach pour confirmer l'aptitude du QTI à mesurer ce qu'il est censé mesurer (sa validité), c'est-à-dire le construit.

Quatorze études parmi les 23 retenues ont précisé le type de fidélité, soit la consistance interne (voir annexe 1). Cette dernière a été mesurée par l'alpha de Cronbach dans tous les cas. Maulana *et al.* (2012) et Kokkinos *et al.* (2009) ont ajouté le coefficient de corrélation intra-classe (CCI) pour justifier la fidélité du QTI. L'alpha de Cronbach a été calculé pour chacune des huit échelles du QTI. Les seuils de recevabilité et leur interprétation diffèrent d'une étude à une autre. Passini *et al.* (2015.) et Maulana *et al.* (2012), par exemple, ont confirmé la fidélité des huit échelles sans justifier leurs décisions, bien que leurs résultats renferment au moins trois échelles ayant des consistances internes inférieures à .70. D'autres comme Sivan et Chan (2013), Lee *et al.* (2003), Kim *et al.* (2000) et Goh et Fraser (1996) ont justifié la fidélité de leur version en se référant à des études antérieures. Ces auteurs ont comparé les seuils des alphas de Cronbach de leurs études avec ceux qui ont été publiés dans le passé. Néanmoins, dans leurs comparaisons, ils n'ont pas pris en considération que le seuil de l'alpha de Cronbach dépend de plusieurs facteurs comme la taille de l'échantillon, le nombre d'items par échelle, etc. (Cho & Kim, 2014).

Dans l'étude de Goh et Fraser (1996), l'alpha de trois échelles du QTI (incertitude, exigence et bienveillance) était inférieur à .60. Pour les autres études, au moins une échelle avait un alpha inférieur à .70. Malgré cela, ces auteurs ont quand même confirmé la fidélité du QTI. Rares sont les auteurs qui ont justifié leurs décisions en se basant sur les seuils d'alpha recommandés par les spécialistes comme Nunnally (1978). D'ailleurs, sur les 23 études analysées, seulement sept d'entre elles se sont référées aux seuils existants pour justifier leurs décisions. Wubbels & Levy (1989) se sont basés sur le seuil de Nunnally (1967). Koul & Fisher (2006), den Brok *et al.* (2003) et Levy *et al.* (2003) ont utilisé ce même seuil bien que Nunnally ait recommandé, après des années de recherche, des seuils plus élevés se situant dans l'intervalle [0.70 - 0.95] (Peterson, 1994). Koul et Fisher (2005, 2006) ont même utilisé le seuil fixé par Cronbach qui remonte au début des années 50 (égales ou supérieures à 0,50). De telles valeurs sont-elles acceptables et fiables de nos jours ?

Parmi les 23 articles analysés, seuls Wubbels et Levy (1989) ont testé la fidélité temporelle. Ces derniers ont confirmé ce type de fidélité en se basant sur le coefficient test-retest (entre 0.65 et 0.84) calculé sur un groupe de 80 élèves, échelonné sur une période de quatre semaines. Dans les autres 22 articles, seuls Lapointe et Legault (1999), Kokkinos *et al.* (2009) et Koul et Fisher (2005) précisent que leur collecte des données a été effectuée après quelques semaines du début de l'année. Les auteurs n'ont pas précisé cette information. À défaut de connaître ce détail, on risque d'avoir des réponses qui ne reflètent pas la réalité puisqu'au début de l'année scolaire les perceptions sont instables (Wubbels et Leavy, 1993).

4. Discussion

La qualité d'une étude scientifique est tributaire de tout un ensemble d'éléments dont les qualités psychométriques des outils de mesure. Ces outils doivent être suffisamment fidèles et valides pour permettre la généralisation des résultats trouvés. Cependant, la documentation et l'expérience ont montré que la validité et la fidélité sont plus facilement applicables pour les mesures des propriétés physiques et physiologiques (fréquence cardiaque, tension artérielle) que pour les mesures des caractéristiques psychologiques (attitudes, comportements, traits de personnalité, opinions...). Une façon d'améliorer la qualité des questionnaires psychométriques consiste à faire la critique de démarches dans le but de formuler des recommandations pour les futures recherches.

Cet article s'inscrit dans ce cadre et a comme objectif de faire l'analyse critique des démarches de vérification des qualités psychométriques du QTI. Les critiques que nous apportons ont été basées essentiellement sur la validité factorielle, structurale et la cohérence interne et temporelle de ce questionnaire.

Le QTI a bénéficié depuis des dizaines d'années d'une reconnaissance internationale (Wubbels et Brekelmans, 2005) et ses dimensions sont, de nos jours, reconnues et acceptées comme des descripteurs universels de l'interaction humaine en général et de l'interaction enseignant-enseigné en particulier (Wubbels et Brekelmans, 2012). Malheureusement, les chercheurs prennent pour acquis qu'il possède des qualités psychométriques suffisantes. Pour preuve, nous n'avons trouvé qu'une seule étude (den Brok *et al.*, 2003a) qui a jeté un regard critique sur le QTI. Cependant, cette dernière n'avait pas pour objectif la critique du QTI; les auteurs ont simplement mentionné l'existence de quelques faiblesses méthodologiques dans les démarches de vérification des qualités psychométriques de ce questionnaire, utilisées dans le passé. Dès lors, nous considérons que nous sommes les premiers à mener une critique approfondie des démarches de validation et de fidélité de ce questionnaire, outil, pour rappel, qui est utilisé pour étudier la relation enseignant-élève, depuis plus de 25 ans (Passini *et al.*, 2015). Notre étude nous a permis, en nous basant sur 23 articles scientifiques publiés dans différentes disciplines scolaires, périodes et contextes, de confirmer la présence de lacunes de plus ou moins grande importance dans chacune des étapes de validation du QTI. Ces lacunes sont susceptibles de remettre en question la crédibilité des résultats et des interprétations éducatives qui résultent de son usage et, par extension, du questionnaire lui-même. Cette constatation converge avec les critiques formulées à titre informatif par den Brok *et al.* (2003a) mais aussi avec les éléments soulevés par des études antérieures qui ont critiqué les démarches de validation d'autres questionnaires dans d'autres domaines (ex. en communication, en psychologie, etc.) (Fabrigar *et al.*, 1997). Parmi ces études, nous pouvons citer celle de Bourque *et al.* (2006) qui a critiqué les démarches de validation des questionnaires utilisés en éducation, domaine de notre étude. En se basant sur 1089 études publiées entre 1995 et 2005, ces auteurs se sont concentrés sur les démarches de validation factorielles pour montrer que « dans la majorité des cas, les chercheurs de divers pays font un usage inadéquat de l'analyse factorielle ou n'en rapportent pas les résultats de façon convenable » (p. 325). D'une façon plus générale, il semble que plusieurs questionnaires psychométriques, spécifiquement ceux traduits en français, sont marqués par plusieurs faiblesses susceptibles de remettre en cause leur validité et leur fidélité (Vallerand, 1989). Nous supposons que l'avènement des logiciels de statistiques qui automatisent et

systématisent le traitement des données sans toutefois les rendre intelligibles et la pression croissante à la publication ont amené les chercheurs à être de moins en moins vigilants face aux démarches de validation adoptées (Beavers *et al.*, 2013).

Revenons au QTI pour rappeler qu'il s'agit d'un questionnaire qui a subi plusieurs modifications au fil du temps, ce qui a donné lieu à plusieurs versions en différentes langues, appliquées sur des catégories d'âge variées, etc. D'ailleurs, on ne trouve pas d'études antérieures ayant montré que les versions adaptées du QTI sont représentatives et même homologues à la version originale. Les chercheurs doivent être conscients des différences qui existent entre ces versions pour choisir celle qui est la mieux adaptée à leur contexte d'étude, chose qui n'est pas toujours respectée. Nous pouvons ici faire appel à Fortin (1994) qui est d'avis qu'un même instrument ou qu'une même échelle de mesure peut être valide dans un contexte et non pas dans un autre et peut avoir une fidélité suffisante à certaines fins plutôt qu'à d'autres. C'est pourquoi il ne suffit pas de s'interroger sur la validité et la fidélité d'un instrument dans l'absolu, mais plutôt de se référer à un contexte défini (Brinberg & McGrath, 1985). Rappelons, par ailleurs, que le QTI est basé sur les perceptions des enseignants et des élèves (Wubbels & Levy, 1989). Or, la perception est un mécanisme de recherche de la représentation la plus claire et la plus précise de la réalité, qui dépend de plusieurs facteurs internes et externes, dans le but d'assurer la survie de l'être humain (Toniolo, 2009). Par ailleurs, la démarche de vérification des qualités psychométriques du QTI comme questionnaire basé sur les perceptions nécessite une bonne définition du contexte d'étude et des aspects socioculturels pour contrôler l'effet des facteurs externes et internes. Cependant, on a remarqué que seulement une minorité des articles analysés en présente une description détaillée.

Les spécialistes en méthodologie de la recherche insistent fortement sur la vérification de quelques prémisses avant d'effectuer des analyses factorielles, structurales et même des démarches de fidélité (Beavers *et al.*, 2013; Bourque *et al.*, 2006; Fabrigar *et al.*, 1997). En ce qui concerne la taille de l'échantillon, nos résultats ont montré que, dans la majorité des cas, il est opportun de choisir des tailles d'échantillons suffisantes pour mener des analyses factorielles. Ceci converge avec ce qui a été soulevé par Bourque *et al.* (2006) quand ils ont critiqué les démarches de validation des différents questionnaires en éducation. Toutefois, les auteurs ne semblent pas justifier la taille de leur échantillon en se référant notamment aux ratios recommandés par Fergusson et Cox (1993) ni à d'autres arguments. Or, selon les spécialistes des AFE, un échantillon de taille inadéquate peut nuire au processus des analyses factorielles en produisant des résultats invalides, aléatoires et non fiables (Beavers *et al.*, 2013 ; Osborne & Costello, 2004 ; Pett *et al.*, 2003).

La technique d'échantillonnage nécessite aussi une révision pour produire des résultats représentatifs de la réalité. En effet, le choix d'un échantillon non probabiliste de convenance présente plus de risque d'introduire des biais systématiques qu'un échantillon probabiliste à caractère aléatoire, particulièrement quand on se base sur les perceptions qui peuvent varier selon plusieurs paramètres (âge, niveau socio-économique, sexe, etc.) (Genoud, 2004).

La matrice de corrélations sur laquelle se basent les analyses factorielles peut biaiser les résultats si elle ne se prête pas à des analyses factorielles exploratoires (Bourque *et al.*, 2006). Il est étonnant que cet élément n'ait été vérifié que dans seulement deux articles sur 23. Ces

informations supplémentaires sont nécessaires à ce propos pour qu'on puisse juger la qualité de la démarche effectuée.

D'après les articles analysés, il semble que, dans la plupart du temps, les chercheurs qui tentent de valider le QTI prennent pour acquis qu'il s'agit d'un questionnaire valide ayant des fondements théoriques solides. Par ailleurs, leurs démarches de vérification de la validité factorielle, structurale ou de la fidélité sont moins rigoureuses que ce qui est recommandé par les spécialistes en méthodologie ou en statistiques. L'analyse des différentes démarches de vérification des qualités psychométriques du QTI nous a permis de soulever plusieurs lacunes dans chacun des articles analysés. Nous n'avons trouvé aucune démarche ayant suivi toutes les recommandations méthodologiques et statistiques nécessaires pour vérifier la validité et la fidélité d'un questionnaire psychométrique. Les résultats de notre étude démontrent la nécessité de cesser de prendre pour acquis que ce questionnaire possède des qualités psychométriques suffisantes. Les chercheurs souhaitant utiliser ce questionnaire dans le futur doivent être conscients de ses lacunes pour mener des démarches de validation rigoureuses. Une telle posture pourra améliorer les qualités psychométriques de ce questionnaire dans le but de mieux étudier la relation enseignant-élève.

En nous basant sur cette étude et sur les études antérieures qui ont critiqué d'autres questionnaires, nous supposons que la validité transculturelle d'un questionnaire n'est jamais assurée. Par conséquent, avant l'administration de n'importe quel questionnaire, l'équipe de recherche doit s'assurer de la pertinence, de la signification et de la sensibilité de chacun des éléments du questionnaire dans le nouveau contexte. Entre autres, il est important de vérifier s'il possède des qualités psychométriques suffisantes pour être utilisé avec la population ou l'échantillon choisi. Ces vérifications nous permettent de produire des résultats qui reflètent la réalité et non pas des résultats biaisés par les qualités psychométriques incertaines des outils de collecte utilisés.

5. Conclusion et recommandations

En éducation, domaine de notre étude, bien que les questionnaires ne cessent de se développer pour documenter ou évaluer un aspect ou un autre de ce domaine (Bourque *et al.*, 2006), la documentation demeure étonnamment pauvre en conseils relatifs aux seuils et aux critères révélant s'il s'agit d'une validité et/ou fidélité suffisante (Peterson, 1994). Face à cela et, peu importe s'il s'agit d'un nouvel instrument ou d'une version traduite dans une autre langue, une attention particulière doit être accordée aux démarches de vérification de la validité et de la fidélité de ces instruments. Les chercheurs doivent suivre des formations et doivent mettre à jour leurs connaissances en méthodologie et en statistiques pour pouvoir effectuer les bons choix et prendre les bonnes décisions. La transposition culturelle et linguistique correcte n'est jamais assurée ; il faut toujours la vérifier. Autrement dit, un questionnaire valide au Canada et en France n'est pas nécessairement valide en Tunisie et le questionnaire valide avec des élèves du secondaire n'est pas nécessairement transposable à ceux du primaire. Dans ces circonstances, en nous basant sur ce qui a été soulevé dans cette étude, nous pouvons émettre des réserves quant à plusieurs démarches adoptées pour établir la validité et la fidélité du QTI. Par conséquent, les résultats de même que les conclusions des travaux publiés à partir de l'une ou l'autre version de cet instrument peuvent être questionnés.

6. Références bibliographiques

- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical assessment, research & evaluation*, 18(6), 1-13.
- Bourque, J., Poulin, N., & Cleaver, A. F. (2006). Évaluation de l'utilisation et de la présentation des résultats d'analyses factorielles et d'analyses en composantes principales en éducation. *Revue des sciences de l'éducation*, 32(2), 325-344.
- Brekelmans, M. (1989). Interpersonal teacher behaviour in the classroom. *Utrecht, The Netherlands: WCC*.
- Brinberg, D., & McGrath, J. E. (1985). Validity and the research process. In *Validity and the Research Process*: Sage Publications.
- Cattell, R. (1966). Validation and intensification of the sixteen personality factor questionnaire. *Readings in Clinical Psychology*, 241-254.
- Cho, E., & Kim, S. (2014). Cronbach's coefficient alpha well known but poorly understood. *Organizational Research Methods*, doi :1094428114555994.
- Cohen, J., McCabe, L., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers college record*, 111(1), 180-213.
- Coll, R. K., Taylor, N., & Fisher, D. L. (2002). An application of the Questionnaire on Teacher Interaction and College and University Classroom Environment Inventory in a multicultural tertiary context. *Research in Science & Technological Education*, 20(2), 165-183.
- Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational research methods*, 6 (2), 147-168.
- Debarbieux, E. (1996). La violence en milieu scolaire: perspectives comparatives portant sur 86 établissements. *Bordeaux: Université de Bordeaux II. e Estado*, 19(1), 21-51.
- Debarbieux, É. (2015). Du "climat scolaire": définitions, effets et politiques publiques. *Éducation et formations* (88-89), 11-27. En ligne: http://cache.media.education.gouv.fr/file/revue_88-89/62/8/depp-2015-EF-88-89-climat-scolaire-definitions-effets-politiques-publiques_510628.pdf
- den Brok, P., Brekelmans, M., Levy, J., & Wubbels, T. (2002). Diagnosing and improving the quality of teachers' interpersonal behaviour. *International journal of educational management*, 16(4), 176-184.
- den Brok, P., Fisher, D., Brekelmans, M., Rickards, T., Wubbels, T., Levy, J., & Waldrup, B. (2003). *Students' Perceptions of Secondary Science Teachers' Interpersonal Style in Six Countries: A Study on the Cross National Validity of the Questionnaire on Teacher Interaction*.
- den Brok, P., Levy, J., Rodriguez, R., & Wubbels, T. (2002a). Perceptions of Asian-American and Hispanic-American teachers and their students on teacher interpersonal communication style. *Teaching and Teacher Education*, 18(4), 447-467.
- den Brok, P., Levy, J., Wubbels, T., & Rodriguez, M. (2003). Cultural influences on students' perceptions of videotaped lessons. *International Journal of Intercultural Relations*, 27(3), 355-374.
- den Brok, P., Fisher, D., Brekelmans, M., Rickards, T., Wubbels, T., Levy, J., & Waldrup, B. (2003a). *Students' perceptions of secondary science teachers' interpersonal style in six countries: A study on the validity of the Questionnaire on Teacher Interaction*. Communication présentée annual meeting of the American Educational Research Association, Chicago.

- Dumas, D. (2009). *Perceptions des comportements interpersonnels d'enseignants et d'enseignantes d'éducation physique et à la santé par des élèves du secondaire*. Mémoire de maîtrise en sciences et en kinanthropologie, Faculté d'éducation physique et sportive, Université de Sherbrooke, Sherbrooke
- Durand, C. (2005). L'analyse factorielle et l'analyse de fidélité. *Notes de cours*. Université de Montréal, département de sociologie, Montréal, Québec.
- Fabrigar, L. R., Visser, P. S., & Browne, M. W. (1997). Conceptual and methodological issues in testing the circumplex structure of data in personality and social psychology. *Personality and Social Psychology Review*, 1(3), 184-203.
- Ferguson, E. & Cox, T. (1993). Exploratory factor analysis: A users' guide. *International Journal of Selection and Assessment*, 1(2), 84-94.
- Field, A. (2009). *Discovering statistics using IBM SPSS statistics*. (3th edition^e éd.). Sage.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. (4th edition^e éd.). Los Angeles : SAGE.: Sage.
- Fisher, D., & Rickards, T. (1998). Associations between teacher-student interpersonal behaviour and student attitude to mathematics. *Mathematics Education Research Journal*, 10(1), 3-15.
- Fisher, D., Fraser, B. J., & Rickards Fortin, F. (1994). Propriétés métrologiques des instruments de mesure (fidélité-validité). *Recherche en Soins Infirmiers*, 39, 58-62.
- Fortin, F. (1994). Propriétés métrologiques des instruments de mesure (fidélité-validité). *Recherche en Soins Infirmiers*, 39, 58-62.
- Fraser, B. J., Aldridge, J. M., & Soerjaningsih, W. (2010). Instructor-student interpersonal interaction and student outcomes at the university level in Indonesia. *The Open Education Journal*, 3(1).
- Genoud, P. A. (2004). *Perception des interactions maître-élèves. L'apport du regard des apprenants sur le profil interactionnel des enseignants en formation*. Thèse de doctorat en lettres et sciences humaines. Suisse: Université de Fribourg.
- Goh, S. C., & Fraser, B. J. (1996). Validation of an elementary school version of the Questionnaire on Teacher Interaction. *Psychological Reports*, 79(2), 515-522.
- Gordon, T. (1979). Enseignants efficaces: enseigner et être soi-même (Trad, par L. B. Lalanne). *Montreal: Editions du Jour Inc*.
- Henderson, D., Fisher, D., & Fraser, B. (2000). Interpersonal behavior, laboratory learning environments, and student outcomes in senior biology classes. *Journal of Research in Science Teaching*, 37(1), 26-43.
- Henson, R. K., Capraro, R. M., & Capraro, M. M. (2001). Reporting Practice and Use of Exploratory Factor Analysis in Educational Research Journals.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185.
- Huot, R. (2003). *Méthodes quantitatives pour les sciences humaines*. Presses Université Laval. Lapointe, J., et Legault, F. (1999). Version francophone du Questionnaire for Teacher Interaction en contexte québécois. *Mesure et évaluation en éducation*, 22(2-3), 1-19.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*.

- Kim, H.-B., Fisher, D. L., & Fraser, B. J. (2000). Classroom environment and teacher interpersonal behaviour in secondary science classes in Korea. *Evaluation & Research in Education*, 14(1), 3-22. DOI: 10.1080/09500790008666958
- Kokkinos, C. M., Charalambous, K., & Davazoglou, A. (2009). Interpersonal teacher behaviour in primary school classrooms: A cross-cultural validation of a Greek translation of the Questionnaire on Teacher Interaction. *Learning Environments Research*, 12(2), 101-114.
- Koul, R. B., & Fisher, D. L. (2005). Cultural background and students' perceptions of science classroom learning environment and teacher interpersonal behaviour in Jammu, India. *Learning Environments Research*, 8(2), 195-211.
- Koul, R. B., & Fisher, D. L. (2006). *Students' perceptions of teachers' interpersonal behaviour and identifying exemplary teachers*. Communication présentée Experience of learning. Proceedings of the 15 th Annual teaching learning forum.
- Lai, H. R., Chou, W. L., Miao, N. F., Wu, Y. P., Lee, P. H., & Jwo, J. C. (2015). A comparison of actual and preferred classroom environments as perceived by middle school students. *Journal of School Health*, 85(6), 388-397.
- Lapointe, J., & Legault, F. (1999). Version francophone du Questionnaire for Teacher Interaction en contexte québécois. *Mesure et évaluation en éducation*, 22(2-3), 1-19.
- Leary, T. F. (1957). *Interpersonal Diagnosis of Personality: A Functional Theory and Methodology for Personality Evaluation*.
- Lee, S. S., Fraser, B. J., & Fisher, D. L. (2003). Teacher–student interactions in Korean high school science classrooms. *International Journal of Science and Mathematics Education*, 1(1), 67-85.
- Légis Québec. (2017). *Régime pédagogique de l'éducation préscolaire, de l'enseignement primaire et de l'enseignement secondaire* (Editeur officiel du Québec), 44. En ligne: <http://legisquebec.gouv.qc.ca/fr/showdoc/cr/I-13.3,%20r.%208#se:18>
- Levy, J., Wubbels, T., Den Brok, P., & Brekelmans, M. (2003). Students' perceptions of interpersonal aspects of the learning environment. *Learning Environments Research*, 6(1), 5-36.
- Maulana, R., Opdenakker, M. C., den Brok, P., & Bosker, R. J. (2012). Teacher–student interpersonal behavior in secondary mathematics classes in Indonesia. *International Journal of Science and Mathematics Education*, 10(1), 21-47.
- Nunnally, J. (1967). Psychometric theory. *New York:McGraw-Hill*, p.194.
- Nunnally, J. (1978). Psychometric theory, 2d ed. *New York:McGraw-Hill*.
- Osborne, J. W., & Costello, A. B. (2009). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pan-Pacific Management Review*, 12(2), 131-146.
- Paillé, P. (2007). La méthodologie de recherche dans un contexte de recherche professionnalisante: douze devis méthodologiques exemplaires. *Recherches qualitatives*, 27(2), 133-151.
- Park, H. S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research*, 28(4), 562-577.
- Passini, S., Molinari, L., & Speltini, G. (2015). A validation of the Questionnaire on Teacher Interaction in Italian secondary school students: the effect of positive relations on motivation and academic achievement. *Social Psychology of Education*, 18(3), 547-559.

- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of consumer research*, 21(2), 381-391.
- Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Sage.
- Pohlmann, J. T. (2004). Use and interpretation of factor analysis in The Journal of Educational Research: 1992-2002. *The Journal of Educational Research*, 98(1), 14-23.
- Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological assessment*, 12(3), 287.
- Rickards, T., & Fisher, D. (2000). *Three Perspectives on Perceptions of Teacher-Student Interaction: A Seed for Change in Science Teaching*.
- Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and social psychology bulletin*, 28(12), 1629-1646.
- Scott, R. H., & Fisher, D. L. (2004). Development, validation and application of a Malay translation of an elementary version of the Questionnaire on Teacher Interaction. *Research in Science Education*, 34(2), 173-194.
- Sivan, A., & Chan, D. W. (2013). Teacher interpersonal behaviour and secondary students' cognitive, affective and moral outcomes in Hong Kong. *Learning Environments Research*, 16(1), 23-36.
- Sivan, A., Chan, D. W., & Kwan, Y. W. (2014). Psychometric Evaluation of the Chinese Version of the Questionnaire on Teacher Interaction (C-QTI) in Hong Kong. *Psychological reports*, 114(3), 823-842.
- Telli, S., den Brok, P., & Cakiroglu, J. (2007). Students' perceptions of science teachers' interpersonal behaviour in secondary schools: Development of a Turkish version of the Questionnaire on Teacher Interaction. *Learning Environments Research*, 10(2), 115-129.
- Toniolo, A.-M. (2009). Le comportement: entre perception et action, un concept à réhabiliter. *L'Année psychologique*, 109(01), 155-193.
- Vallerand, R. J. (1989). Vers une méthodologie de validation trans-culturelle de questionnaires psychologiques: Implications pour la recherche en langue française. *Canadian Psychology/Psychologie Canadienne*, 30(4), 662.
- Watzlawick, P., Bavelas, J. B., Jackson, D. D., & O'Hanlon, B. (2011). *Pragmatics of human communication: A study of interactional patterns, pathologies and paradoxes*. WW Norton & Company.
- Wei, M., Den Brok, P., & Zhou, Y. (2009). Teacher interpersonal behaviour and student achievement in English as a Foreign Language classrooms in China. *Learning Environments Research*, 12(3), 157-174.
- Wilkins, J. (2014). The development of a scale to explore the multidimensional components of good student-teacher relationships. *Education Research and Perspectives (Online)*, 41, 154.
- Wubbels, T. (1985). *Discipline Problems of Beginning Teachers, Interactional Teacher Behaviour Mapped Out*.
- Wubbels, T., & Brekelmans, M. (2005). Two decades of research on teacher-student relationships in class. *International Journal of Educational Research*, 43(1), 6-24.
- Wubbels, T., Creton, H., & Houymayers, H. (1985). *Discipline Problems of Beginning Teachers, Interactional Teacher Behaviour Mapped Out*.

Wubbels, T., & Levy, J. (1989). *A Comparison of Dutch and American Interpersonal Teacher Behavior*. En ligne: <https://eric.ed.gov/?id=ED307311>.

Wubbels, T., & Levy, J. (1993). *Do you know what you look like?: Interpersonal relationships in education*. London, Washington, DC: The Falmer Press.

Annexe 1 : références (articles analysés seulement)

1. Coll, R. K., Taylor, N., & Fisher, D. L. (2002). An application of the questionnaire on teacher interaction and college and university classroom environment inventory in a multicultural tertiary context. *Research in Science & Technological Education*, 20(2), 165-183. Doi :10.1080/0263514022000030435.
2. den Brok, P., Levy, J., Wubbels, T., & Rodriguez, M. (2003). Cultural influences on students' perceptions of video taped lessons. *International Journal of Intercultural Relations*, 27(3), 355-374.
3. Fisher, D., Fraser, B. J., & Rickards, T. W. (1997). *Gender and Cultural Differences in Teacher-Student Interpersonal Behavior* (Reports - Research; Speeches/Meeting Papers). Article présenté lors « the Annual Meeting of the American Educational Research Association », 24-28, Chicago, IL.
4. Fraser, B. J., Aldridge, J. M., & Soerjaningsih, W. (2010). Instructor-student interpersonal interaction and student outcomes at the university level in Indonesia. *The Open Education Journal*, 3(1).
5. Goh, S. C. & Fraser, B. J. (1996). Validation of an elementary school version of the Questionnaire on Teacher Interaction. *Psychological Reports*, 79(2), 515-522.
6. Henderson, D., Fisher, D., & Fraser, B. (2000). Interpersonal behavior, laboratory learning environments, and student outcomes in senior biology classes. *Journal of Research in Science Teaching*, 37(1), 26-43.
7. Kim, H. B., Fisher, D. L., & Fraser, B. J. (2000). Classroom environment and teacher interpersonal behaviour in secondary science classes in Korea. *Evaluation & Research in Education*, 14(1), 3-22.
8. Kokkinos, C. M., Charalambous, K. & Davazoglou, A. (2009). Interpersonal teacher behaviour in primary school classrooms: A cross-cultural validation of a Greek translation of the Questionnaire on Teacher Interaction. *Learning Environments Research*, 12(2), 101-114.
9. Koul, R. B. & Fisher, D. L. (2005). Cultural background and students' perceptions of science classroom learning environment and teacher interpersonal behaviour in Jammu, India. *Learning Environments Research*, 8(2), 195-211.
10. Koul, R. B. & Fisher, D. L. (2006). Students' perceptions of teachers' interpersonal behaviour and identifying exemplary teachers. In *Experience of learning. Proceedings of the 15th Annual teaching learning forum* (pp. 1-2).
11. Lapointe, J. & Legault, F. (1999). Version francophone du Questionnaire for Teacher Interaction en contexte québécois. *Mesure et évaluation en éducation*, 22(2-3), 1-19.
12. Lee, S. S., Fraser, B. J. & Fisher, D. L. (2003). Teacher–student interactions in Korean high school science classrooms. *International Journal of Science and Mathematics Education*, 1(1), 67-85.
13. Levy, J., Wubbels, T., Den Brok, P. & Brekelmans, M. (2003). Students' perceptions of interpersonal aspects of the learning environment. *Learning Environments Research*, 6(1), 5-36.
14. Maulana, R., Opendakker, M. C., den Brok, P. & Bosker, R. J. (2012). Teacher–student interpersonal behavior in secondary mathematics classes in Indonesia. *International Journal of Science and Mathematics Education*, 10(1), 21-47.
15. Passini, S., Molinari, L. & Speltini, G. (2015). A validation of the Questionnaire on Teacher Interaction in Italian secondary school students: the effect of positive relations on motivation and academic achievement. *Social Psychology of Education*, 18(3), 547-559.

16. Rickards, T. & Fisher, D. (2000). Three Perspectives on Perceptions of Teacher-Student Interaction: A Seed for Change in Science Teaching.
17. Scott, R. H. & Fisher, D. L. (2004). Development, validation and application of a Malay translation of an elementary version of the Questionnaire on Teacher Interaction. *Research in Science Education*, 34(2), 173-194.
18. Sivan, A. & Chan, D. W. (2013). Teacher interpersonal behaviour and secondary students' cognitive, affective and moral outcomes in Hong Kong. *Learning Environments Research*, 16(1), 23-36.
19. Sivan, A. Chan, D. W. & Kwan, Y. W. (2014). Psychometric evaluation of the Chinese Version of the Questionnaire on Teacher Interaction (C—QTI) in Hong Kong. *Psychological reports*, 114(3), 823-842.
20. Telli, S., den Brok, P. & Cakiroglu, J. (2007). Students' perceptions of science teachers' interpersonal behaviour in secondary schools: Development of a Turkish version of the Questionnaire on Teacher Interaction. *Learning Environments Research*, 10(2), 115-129.
21. Wei, M., den Brok, P. & Zhou, Y. (2009). Teacher interpersonal behaviour and student achievement in English as a Foreign Language classrooms in China. *Learning Environments Research*, 12(3), 157-174.
22. Wubbels, T. (1985). *Discipline Problems of Beginning Teachers, Interactional Teacher Behaviour Mapped Out*.
23. Wubbels, T., Creton, H. & Houymayers, H. (1985). *Discipline Problems of Beginning Teachers, Interactional Teacher Behaviour Mapped Out*.

Annexe 2 : Présence des principaux critères des AFE dans les 23 articles analysés

Critères	Articles numérotés selon l'annexe 1																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Taille d'échantillon																							
n < 100																							
100 ≤ n < 200																							
200 ≤ n < 300																							
n > 300																							
Matrice de corrélation																							
Vérification des corrélations																							
Déterminant rapporté																							
Indice KMO																							
Test de Bartlett																							
Validité factorielle																							
Méthode d'extraction précisée																							
Critère d'extraction précisé																							
Complexité des chargements vérifiée																							
Méthode de rotation des axes précisée																							
Validité structurale																							
Tableau de corrélation complet																							
Justification de la validité																							
Fidélité																							
Type de fidélité précisé (Consistance interne)																							
CCI																							
Seuil minimum d'alpha de Cronbach accepté																							
< 0,5																							
]0,5; 0,6]																							
]0,6; 0,7]																							
> 0,7																							

Légende :
 (n) : taille d'échantillon ; (+) : critère respecté ; () : critère non respecté.