



# PLAN

<b>PLAN</b> .....	<b>1</b>
<b>Liste des figures</b> .....	<b>5</b>
<b>Liste des tableaux</b> .....	<b>5</b>
<b>INTRODUCTION</b> .....	<b>6</b>
<b>GÉNÉRALITÉS</b> .....	<b>10</b>
A. C'est quoi un environnement de formation .....	11
B. Comment mesurer l'environnement de formation .....	14
1. Différents Instruments mesurant l'environnement de formation .....	15
2. Le Dutch Residency Educational Climate Test (D-RECT) .....	20
<b>MATÉRIEL ET MÉTHODES</b> .....	<b>30</b>
A. Élaboration du questionnaire D-RECT en français.....	31
1. Traduction.....	31
2. Prétest.....	31
B. Distribution du questionnaire .....	34
C. Administration du questionnaire .....	35
D. Méthodes statistiques.....	35
1. Statistiques descriptives.....	35
2. Analyse de la fidélité.....	36
a. Consistance interne .....	36
b. Test —Retest .....	37
3. Evaluation de la validité du construit .....	37
a. Structure du construit : analyse factorielle confirmatoire .....	38
b. Validité convergente .....	43
c. Validité discriminatoire.....	43

4. Logiciel statistique .....	43
<b>RÉSULTATS .....</b>	<b>44</b>
A. Description de la population .....	45
1. Profil démographique .....	45
2. Profil professionnel.....	45
B. Résultats des scores D-RECT.....	47
C. Analyse de la fidélité.....	52
1. Consistance interne .....	52
2. Test-retest .....	52
D. Analyse factorielle confirmatoire .....	53
1. Validité du construit .....	55
2. Validité convergente .....	58
3. Validité discriminatoire .....	58
E. Ajustement du modèle .....	59
<b>DISCUSSION .....</b>	<b>61</b>
A. Méthodologie de validation .....	62
1. Traduction.....	63
2. L'évaluation et la modification de la version préliminaire.....	64
3. Le prétest .....	67
4. Évaluation de la validité concomitante et de contenu .....	67
5. Administration du questionnaire .....	68
6. Analyse de la fidélité.....	69
7. Validité du construit .....	71
B. Points forts .....	79

C. Limites du travail .....	80
D. Implications et futures pistes de recherche .....	81
<b>CONCLUSION .....</b>	<b>82</b>
<b>RÉSUMÉS.....</b>	<b>85</b>
<b>BIBLIOGRAPHIE.....</b>	<b>91</b>

## Liste des figures

**Figure 1** : Instruments de mesure de l'environnement de formation

**Figure 2** : Éléments du questionnaire Postgraduate Hospital Educational Environment Measure (PHEEM) et sous-échelles correspondantes

**Figure 3** : Première version du D-RECT par Boor et al

**Figure 4** : Dutch Residency Education Climate Test (D-RECT). Version anglaise

**Figure 5** : Diagramme de l'analyse factorielle confirmatoire.

## Liste des tableaux

**Tableau 1** : Version française du Dutch Residency Educational Climate Test

**Tableau 2** : Caractéristiques démographiques de la population

**Tableau 3** : Résultats des sous-domaines du D-RECT.

**Tableau 4** : Scores moyens des questions du questionnaire D-RECT

**Tableau 5** : Analyse de la distribution par les tests d'asymétrie et de Kurtosis.

**Tableau 6** : Test de Cronbach alpha pour les dimensions du score D-RECT.

**Tableau 7** : Coefficients de régression de l'analyse factorielle confirmatoire.

**Tableau 8** : Indicateurs de la validité du construit de l'analyse factorielle confirmatoire

**Tableau 9** : Étude de la validité convergente

**Tableau 10** : Indicateurs de validité du construit du modèle original et du modèle modifié

**Tableau 11** : Résultats de la consistance interne comparés à la littérature

# INTRODUCTION

La formation médicale post-graduée se déroule dans la plupart des pays dans le cadre de système du résidanat. L'objectif de cette formation est de produire des médecins spécialistes compétents capables de répondre aux besoins de la population et du pays. La qualité de formation des résidents est donc un enjeu majeur pour les décideurs et les responsables pédagogiques dans les facultés de médecine et les services cliniques universitaires.

Parmi les éléments importants des systèmes de résidanat, le concept d'environnement, ou climat, de formation prend une place de plus en plus importante dans la recherche en pédagogie(1).

Les résidents reçoivent leur formation principalement en milieu de travail clinique. Ils perçoivent un certain climat d'apprentissage lorsqu'ils travaillent dans leurs départements. Leur impression de ce climat peut être « bonne », « inspirante » ou même « optimale », mais il est tout aussi probable qu'elle soit « mauvaise », « décourageante » ou même « déplorable », avec toutes les nuances possibles d'appréciation intermédiaires. L'important est que le climat d'apprentissage local façonne de manière cruciale les expériences d'apprentissage dans le contexte hospitalier. Les résidents apprennent et travaillent dans un état de mutation constante. C'est avant tout inhérent au travail dans la réalité complexe des services hospitaliers, mais c'est aussi lié à l'évolution des programmes d'études et de la gouvernance.



Les climats d'apprentissage nous renseignent sur les contextes dans lesquels les résidents évoluent. C'est un concept qui concerne les multiples facettes de la formation des résidents. Il reflète la façon dont les individus abordent l'apprentissage dans les services cliniques et intègre des perceptions communes de ces personnes sur des thèmes tels que l'atmosphère, la supervision et l'état de l'apprentissage. Les climats d'apprentissage sont construits à travers les interactions des apprenants et des autres travailleurs de la santé et sont influencés par les arrangements organisationnels(1).

La mesure des climats d'apprentissage peut servir d'indicateur général du fonctionnement de l'éducation d'un département en raison de la polyvalence de sa construction.

Il existe plusieurs instruments pour évaluer l'environnement de formation dans le contexte du résidanat. Mais ces instruments n'ont pas été validés dans le contexte du résidanat au Maroc, et donc, nous ne pouvons évaluer l'environnement de formation chez nous de manière fiable.

Le Dutch Residency Educational Climate Test (D-RECT) est un instrument valide fiable pour évaluer l'environnement de formation dans le contexte du résidanat. Il a été créé par une méthodologie rigoureuse et il est donc l'instrument le plus solide rapporté dans la littérature.

L'objectif principal de notre travail est de valider la traduction française du D-RECT dans le contexte du résidanat au Maroc afin qu'il puisse être utilisé pour l'évaluation de l'environnement de formation au Maroc.

La validation psychométrique est une procédure rigoureuse et longue. À cause de sa complexité, nous nous sommes aussi fixés comme objectif secondaire de fournir aux futurs chercheurs une description détaillée de la procédure de validation avec un lien aux principales références publiées.

Dans un premier temps, nous présentons une revue de la littérature pour expliquer ce qu'est un environnement de formation et son importance. Ensuite, nous détaillerons la méthodologie de traduction et validation du questionnaire D-RECT en français.

# GÉNÉRALITÉS

## **A. C'est quoi un environnement de formation ?**

L'environnement de formation est un concept important et large, qui ne concerne pas uniquement les résidents, mais tous les étudiants en médecine. Les termes environnement et climat sont souvent utilisés de manière interchangeable dans la littérature pédagogique. Genn(2) définit le terme climat comme l'environnement éducatif d'un établissement tel qu'il est perçu par les étudiants. Dans l'étude de Rothman&Ayoade(3), l'environnement d'apprentissage est défini comme ce que les étudiants perçoivent et, par conséquent, aucune distinction n'est faite entre les termes environnement et climat. Genn(4) fournit une définition de l'environnement éducatif en tant que «la manifestation d'un programme d'enseignement». Rothman et Ayoade(5) considèrent également l'environnement d'apprentissage comme «une manifestation des effets sur les étudiants des différentes parties du programme». Maudsley (2001) déclare qu'«un environnement d'apprentissage existe partout et à chaque fois que les élèves se rassemblent...», qu'il inclut divers facteurs qui contribuent à une éducation efficace et deviennent l'arrière-plan du programme d'études. Pace et Stern(5,6) suggèrent qu'un environnement éducatif est caractérisé par les pressions, les contraintes, les pratiques, les politiques, les récompenses et les valeurs de la classe ou de l'école, qu'ils qualifient de «presse environnementale». En outre, l'une des questions importantes à se poser lors de la planification d'un cours ou d'un curriculum est la suivante : «Quel environnement ou quel climat d'éducation faut-il favoriser?»(7). Celui qui fait la planification du parcours ou du curriculum doit évaluer les domaines mis en avant et encouragés dans l'institution.

L'environnement de formation à la faculté de médecine fait partie de l'un des facteurs déterminants du comportement de l'étudiant en médecine(4). Les effets d'un environnement éducatif sont substantiels, réels et influents(7). Le climat contribue de manière significative à la prévision du rendement, de la satisfaction et du succès des étudiants(2). La relation entre l'environnement éducatif tel qu'il est perçu ou vécu par les étudiants et leur accomplissement, leur satisfaction et leur succès à la faculté de médecine a fait prendre conscience de la nécessité d'étudier ou d'évaluer l'environnement des études médicales. La Fédération mondiale de l'éducation médicale a considéré l'environnement d'apprentissage comme l'un des objectifs de l'évaluation des programmes d'éducation médicale. Sur la base de tout ce qui précède, une étude de l'environnement devrait faire partie intégrante de l'étude du programme d'études d'une école de médecine.

Dans le cadre spécifique du résidanat(8), le climat de formation peut être conceptualisé comme la perception des résidents des aspects formels et informels de l'éducation (9), y compris la perception de l'atmosphère générale(4) ainsi que des politiques, pratiques et procédures en vigueur dans l'hôpital universitaire(10).

Le climat de formation a été divisé par certains auteurs en trois parties, il s'agissait de «l'environnement physique» (sécurité, nourriture, abri, confort et autres installations), du «climat émotionnel» (sécurité, réactions constructives, soutien et absence d'intimidation et de harcèlement) et du «climat intellectuel» (apprendre avec patients, pertinence pour la pratique, participation active des apprenants basée sur des preuves, motivation et planification(11)).

Au cours de leur formation initiale, les futurs médecins sont rapidement exposés au milieu clinique. Lorsqu'ils deviennent résidents, la majorité de leur temps d'apprentissage se déroule dans les hôpitaux ou les cliniques. La multiplicité des milieux de formation et la variabilité des conditions et des ressources d'un milieu à l'autre engendrent des disparités du contenu éducatif, de l'exposition clinique, ainsi que de l'environnement social et physique(4). C'est la perception de l'environnement par les résidents, et non l'environnement lui-même, qui détermine leurs comportements(4). Les perceptions des résidents sur l'environnement du service ont été utilisées pour prédire leurs résultats. Les résidents sont ceux qui expérimentent l'environnement et c'est leur perception de l'environnement qui devient alors l'un des déterminants de leur comportement(4).

Le processus d'apprentissage requiert un environnement d'apprentissage adéquat, qui doit pouvoir être évalué de manière régulière. Cette évaluation doit particulièrement être réalisée lorsque des changements du curriculum sont effectués. Les gouvernements, les régulateurs et les conseils de formation médicale s'attendent à une formation de haute qualité en résidanat et l'inspectent régulièrement(4). Un moyen d'évaluer la qualité des programmes de formation consiste à évaluer les climats d'apprentissage. Les climats d'apprentissage nous renseignent sur les contextes dans lesquels les résidents évoluent. Les climats d'apprentissage sont construits à travers les interactions des apprenants et des autres travailleurs de la santé et sont influencés par les arrangements organisationnels(1,4).

Il a été montré que lorsque les résidents perçoivent leur climat d'apprentissage comme positif, ils ont davantage tendance à mobiliser leurs connaissances théoriques et des styles d'apprentissages efficaces. En outre, un climat d'apprentissage positif contribue à prévenir l'épuisement professionnel des résidents(1,4,12) et peut notamment favoriser la satisfaction professionnelle et le développement de l'identité professionnelle des résidents.

Les transformations récentes de la formation médicale renforcent la nécessité de pouvoir disposer d'un outil d'évaluation robuste et pratique. L'environnement éducatif constitue une partie intégrante du curriculum et constitue un élément central du climat d'apprentissage(4). Cette dernière notion correspond à un construit théorique qui peut être mesuré spécifiquement par certains instruments.

## **B. Comment mesurer l'environnement de formation ?**

Une mesure quantitative de l'environnement éducatif nécessite l'utilisation d'un inventaire ou d'un instrument. Le choix d'un tel instrument devrait être fondé sur les qualités du processus de mesure utilisant un instrument particulier ou sur la capacité de cet instrument à mesurer l'environnement éducatif. Les qualités, ou souvent appelées caractéristiques psychométriques, relèvent généralement de deux rubriques principales, à savoir la validité et la fiabilité. Un outil d'environnement d'apprentissage valide et fiable permet une mesure significative de l'environnement d'apprentissage d'une institution et permet ainsi de prendre les mesures appropriées pour améliorer l'environnement. Les différences entre les différents contextes éducatifs peuvent nécessiter un inventaire différent des environnements éducatifs, qui convient à la situation spécifique d'une institution donnée.

Dans tout processus de mesure, il est nécessaire de déterminer si l'instrument ou l'inventaire mesure ce qu'il est supposé mesurer, et la validité traite de cette question spécifique. La validité du contenu s'entend de la mesure dans laquelle l'instrument ou l'inventaire mesure le contenu souhaité de la matière(13). La validité de construction peut être définie comme la mesure dans laquelle le test peut être utilisé pour mesurer certaines constructions psychologiques et les résultats interprétés en termes de ces constructions(13). La fiabilité fait référence à la reproductibilité des résultats de mesure ou d'évaluation. Un résultat peut être fiable sur différentes périodes, sur différents évaluateurs ou sur différents échantillons de questions. Un résultat de mesure peu fiable ou incohérent ne peut permettre une interprétation valide du résultat. Une forte corrélation entre les scores des éléments individuels, autrement dit une cohérence interne élevée, indiquerait que les scores mesurent un seul concept. Le degré de fiabilité adéquat pour une mesure donnée dépend : de l'objectif de la mesure, de l'utilisation des résultats, de l'importance de la décision qui sera prise et, finalement, des conséquences résultant de la mesure(13,14).

## **1. Différents Instruments mesurant l'environnement de formation**

Plusieurs instruments ont été développés pour mesurer l'environnement de formation dans le contexte du résidanat de médecine. Ils ont été résumés dans une revue systématique menée par Soemantri et al.(15). L'évaluation systématique a abouti à 178 études considérées comme pouvant faire l'objet d'un examen ultérieur. Des critères prédéfinis ont été appliqués à ces études et enfin, 79 études ont été incluses. Un nombre important d'instruments de mesure de l'environnement éducatif (31 instruments) ont été extraits des études incluses. La plupart de ses instruments ont été développés pour une utilisation dans des domaines spécialisés spécifiques. Les programmes d'enseignement médical post-gradué comprenaient de nombreuses spécialités et niveaux de formation. Chaque spécialité avait sa propre spécificité en



termes d'environnement de formation. Cette situation nécessitait souvent un instrument spécifique pouvant traiter certains aspects de l'environnement de formation dans le contexte particulier. Dans cette étude, sept instruments ont été identifiés pour l'enseignement médical post-gradué ou le résidanat. Nous allons présenter brièvement les principaux instruments décrits dans la littérature avant de décrire le Dutch Residency Educational Climate Test.

No.	Instrument/Inventory	Setting
1	Medical school learning environment Survey – modified (MSLES)	Medicine
2	Learning environment questionnaire (LEQ)	Medicine
3	Questionnaire from Parry et al.	Medicine
4	Veteran affairs (VA) learners' perceptions survey	Medicine (postgraduate, PG)
5	Learning environment assessment (LEA)	Medicine (PG)
6	Medical school environment questionnaire (MSEQ)	Medicine
7	Questionnaire from Robins et al.	Medicine
8	Medical school environment inventory (MSEI)	Medicine
9	Questionnaire from Rotem, Godwin and Du	Medicine (PG)
10	Course valuing inventory (CVI)	Medicine
11	DREEM	Medicine (undergraduate and PG), nursing, dental and chiropractic
12	Operating room educational environment measure (OREEM)	Medicine (PG)
13	Instrument from Pololi and Price	Medicine
14	Surgical theatre educational environment measure (STEEM)	Medicine (undergraduate and PG)
15	Anaesthetic theatre educational environment measure (ATEEM)	Medicine (PG)
16	Practice-based educational environment measure	Medicine (PG)
17	Postgraduate hospital educational environment measure (PHEEM)	Medicine (PG)
18	Questionnaire from Patel and Dauphinee	Medicine
19	Learning Climate Measure from Wangsaturaka	Medicine
20	Questionnaire from Orton – Modified	Nursing
21	Questionnaire from Hart and Rotem	Nursing
22	Clinical learning environment scale (CLE)	Nursing
23	Clinical learning environment inventory (CLEI)	Nursing
24	College and university classroom environment inventory (CUCEI)	Nursing
25	Clinical learning environment and supervision (CLES)	Nursing
26	Clinical post-conference learning environment survey (CPCLES)	Nursing
27	Postgraduate research experience questionnaire (PREQ)	Nursing
28	Questionnaire from Gerzina, McLean and Fairley	Dentistry
29	Clinical education instructional quality (ClinEd IQ)	Dentistry
30	Learning environment survey (LES)	Dentistry
31	Dental student learning environment survey (DSLES)	Dentistry

**Figure 1 : Instruments de mesure de l'environnement de formation(15)**

Le «Surgical Theatre Educational Environment Measure» (STEEM) a été développé en 2004 par Cassar en Écosse. C'est un questionnaire de 40 questions qui a comme objectif de mesurer l'environnement de formation en salle opératoire(15,16). De manière similaire, l'«Operating room educational environment measure» OREEM a été développé au Canada par Kanashiro et al. en 2006 pour l'évaluation de la formation dans les salles d'opération(17).

L'«Anesthetic theater educational environmentmeasure (ATEEM)»a été développé par Holt et al. en 2004 pour l'évaluation de la formation en salle d'anesthésie(18). The practice-based educational environment measure(19) pour l'évaluation de la formation en médecine générale.

Ces quatre instruments sont spécialisés et ne peuvent s'appliquer pour mesurer l'environnement de formation chez tous les résidents.

Le «Learning environment assessment» (LEA) a été développé à la Wayne State University (USA). La validité de contenu de la LEA n'a pas été établie, bien qu'elle ait montré une cohérence interne assez élevée(20).

The Veteran affairs (VA) Learners Perceptions Survey a été conçue pour évaluer la formation des résidents de différentes spécialités dans le cadre de la formation clinique aux États-Unis pour la VA(21).

Le «Dundee Ready Education Environment Measure» (DREEM) a été développé en Écosse pour évaluer l'environnement de formation chez les étudiants en médecine, médecine dentaire, résidanat et les infirmiers en formation(22). Ses points forts étaient sa grande cohérence interne et sa validité concurrente. Cependant, le contenu et les validités de construction n'ont pas été établis, car l'analyse factorielle a montré que certaines modifications étaient nécessaires pour que l'instrument soit appliqué pour l'évaluation de la formation post-graduée.

Le questionnaire «Postgraduate Hospital Educational Environment Measure»(PHEEM)(23) a été développé par Roff et al.(11) pour mieux évaluer les milieux d'apprentissage et d'enseignement des médecins en formation du Royaume-Uni. Ce questionnaire inclut 40 éléments. Les aspects principaux mesurés sont divisés en trois catégories : la perception de l'autonomie, la perception de l'enseignement et la perception du soutien social. Chacun des éléments est coté selon une échelle de Likert à cinq niveaux (de 0 — fortement en désaccord — à 4 — fortement en accord), pour un score maximal de 160. Un guide publié avec l'article original propose des repères pour interpréter le score final (0-40 points : très mauvais; 41-80 points : plusieurs problèmes; 81-120 points : davantage d'éléments positifs que négatifs, mais avec des points à améliorer; 121-160 : excellent). Le PHEEM a été traduit et validé dans plusieurs langues et dans d'autres milieux(24-27) y compris le Français.

Il a également été démontré qu'en utilisant PHEEM, on pouvait obtenir des résultats fiables avec une taille d'échantillon réalisable. De plus, le PHEEM avait été administré à plusieurs groupes d'échantillons, tels que les hauts responsables de l'administration et les bureaux d'enregistrement spécialisés dans différentes spécialités, et présentait des coefficients de fiabilité presque similaires dans ces groupes(24-27). Le PHEEM a été considéré comme approprié pour une utilisation dans l'évaluation de la formation médicale post-graduée en raison de la validité de son contenu, de sa grande fiabilité et de son aptitude à être utilisé dans différents contextes post-gradués. Cependant, certains chercheurs ont mené une analyse factorielle sur le PHEEM et ont montré des résultats très contradictoires en ce qui concerne les facteurs mesurés par cet instrument, ce qui fait que sa répartition en sous-domaines est aujourd'hui controversée.

Sous-échelle	Item
Perception de l'autonomie	1. J'ai un contrat d'embauche qui m'informe sur les horaires de travail.
	4. J'ai eu une séance d'information à mon programme de résidence au début de mon stage.
	5. J'ai des responsabilités appropriées pour mon niveau de formation dans ce stage.
	8. Je dois réaliser des tâches inappropriées pour mon niveau de formation.
	9. Il y a un manuel d'information sur le stage à l'intention des résidents.
	11. Je me fais signaler de façon inappropriée.
	14. Il existe des protocoles de soins clairs dans ce stage.
	17. Mon horaire de travail respecte l'entente collective en vigueur.
	18. J'ai l'opportunité d'offrir une continuité de soins aux patients.
	29. Je sens que je fais partie d'une équipe de travail.
	30. J'ai la possibilité d'acquérir les techniques appropriées à mon niveau de formation.
	32. Ma charge de travail dans ce stage est adéquate.
	34. La formation dans ce stage me fait sentir prêt à être un médecin clinicien autonome.
40. Les superviseurs cliniques encouragent le respect mutuel.	
La perception de l'enseignement	2. Mes superviseurs cliniques établissent des attentes claires.
	3. J'ai du temps privilégié pour l'étude dans ce stage.
	6. J'ai une bonne supervision clinique en tout temps.
	10. Mes superviseurs cliniques ont de bonnes habiletés de communication.
	12. J'ai la possibilité de participer activement à des activités éducatives.
	15. Mes superviseurs cliniques sont enthousiastes.
	21. J'ai accès à un programme éducatif pertinent à mes besoins.
	22. Je reçois régulièrement de la rétroaction par mes superviseurs cliniques.
	23. Mes superviseurs cliniques sont bien organisés.
	27. J'ai suffisamment d'opportunités d'apprentissage clinique pour mes besoins.
	28. Mes superviseurs cliniques sont de bons professeurs.
	31. Mes superviseurs cliniques sont accessibles.
	33. Mes superviseurs cliniques utilisent efficacement les occasions pour enseigner.
37. Mes superviseurs cliniques m'encouragent à apprendre par moi-même.	
39. Les superviseurs cliniques me fournissent une bonne rétroaction sur mes forces et faiblesses.	
La perception	7. Il y a du racisme dans ce stage.
	13. Il y a de la discrimination sexuelle dans ce stage.

du soutien social	16. J'ai une bonne collaboration avec les autres résidents de mon niveau.
	19. J'ai un accès adéquat à des conseils sur la planification de carrière.
	20. L'hôpital a des installations de bonne qualité pour les résidents, surtout lorsqu'ils sont de garde.
	24. Je me sens physiquement en sécurité dans l'environnement hospitalier.
	25. Il existe une culture de non-blâme dans ce stage.
	26. Il y a des services alimentaires adéquats lorsque je suis de garde.
	35. Mes superviseurs cliniques ont de bonnes aptitudes comme mentors.
	36. Je retire beaucoup de plaisir de mon stage actuel.
	38. Il existe des services de soutien/orientation pour les résidents qui échouent leur stage.

**Figure 2 : Éléments du questionnaire Postgraduate Hospital Educational Environment Measure (PHEEM) et sous-échelles correspondantes(23)**

## **2. Le Dutch Residency Educational Climate Test (D-RECT)**

Le D-RECT a été développé et validé aux Pays-Bas en 2011 afin de répondre aux «défauts» des autres instruments sus-cités. Nous allons exposer les étapes qui ont abouti à la version actuelle du D-RECT.

### ➤ **Première version**

Boor et al. a remarqué que les deux instruments les plus utilisés pour mesurer les environnements d'apprentissage (PHEEM, DREEM), manquent d'une base théorique clairement décrite, et leur structure factorielle sous-jacente est contestée(28). Par exemple, PHEEM a été décrit comme possédant trois sous-échelles différentes dans trois publications différentes(26,28-30). Une telle structure de sous-échelles controversée empêche toute possibilité d'avoir une bonne évaluation de l'environnement post-gradué(28).

Pour surmonter les contraintes décrites ci-dessus, Boor and .al ont cherché à développer un nouveau test psychométrique on se basant sur des résultats de recherche qualitative(28). La première version du D-RECT comprenait 83 éléments et 21 sous-échelles. Une équipe composée de quatre résidents, deux tuteurs spécialisés et trois éducateurs en médecine ont vérifié la validité de ses éléments, et ont suggéré des modifications qui ont conduit à un D-RECT préliminaire de 75 items. Pour chaque item, des experts ont été invités à s'accorder sur une échelle de Likert en cinq points (1= totalement en désaccord et 5=totalement d'accord) sur sa pertinence. Une procédure Delphi modifiée parmi les experts a déterminé l'inclusion et l'exclusion finales des items. Pour la validation, le questionnaire a été administré à tous les résidents des Pays-Bas en 2008. Des analyses psychométriques approfondies ont révélé un questionnaire multifactoriel. Une version D-RECT finale comprenant 50 items fut donc créée (Figure 3).

La théorie de la généralisation a été utilisée pour déterminer le nombre de répondants nécessaires à l'obtention de résultats de tests fiables. Les analyses ont abouti à de nombreux tableaux par sous-échelle, qui ont montré que pour un département, un nombre minimum de 11 résidents est nécessaire pour obtenir un résultat fiable, bien que la plupart des sous-échelles puissent être jugées de manière fiable en utilisant les informations fournies par huit résidents. Pour les groupes de départements, quatre résidents de six départements sont nécessaires pour obtenir des résultats de test fiables.

<b>Supervision</b>
1. The guidelines clearly outline when to request input from a supervisor
2. The amount of supervision I receive is appropriate for my level of experience
3. It is clear which attending supervises me
<b>Coaching and assessment</b>
4. I am asked on a regular basis to provide a rationale for my management decisions and actions
5. My attendings coach me on how to communicate with difficult patients
6. My attendings take the initiative to explain their actions
7. My attendings take the initiative to evaluate my performance
8. My attendings take the initiative to evaluate difficult situations I have been involved in
9. My attendings evaluate whether my performance in patient care is commensurate with my level of training
10. My attendings occasionally observe me taking a history
11. My attendings assess not only my medical expertise but also other skills such as teamwork, organization or professional behavior
<b>Feedback</b>
12. My attendings give regular feedback on my strengths and weaknesses
13. Observation forms (i.e., Mini-CEX) are used to structure feedback
14. Observation forms (i.e., Mini-CEX) are used periodically to monitor my progress
<b>Teamwork</b>
15. Attendings, nursing staff, other allied health professionals and residents work together as a team
16. Nursing staff and other allied health professionals make a positive contribution to my training
17. Nursing staff and other allied health professionals are willing to reflect with me on the delivery of patient care
18. Teamwork is an integral part of my training
<b>Peer collaboration</b>
19. Residents work well together
20. Residents, as a group, make sure the day's work gets done
21. Within our group of residents, it is easy to find someone to cover or exchange a call
<b>Professional relations between attendings</b>
22. Continuity of care is not affected by differences of opinion between attendings
23. Differences of opinion between attendings about patient management are discussed in such a manner that is instructive to others present
24. Differences of opinion are not such that they have a negative impact on the work climate
<b>Work is adapted to residents' competence</b>
25. The work I am doing is commensurate with my level of experience
26. The work I am doing suits my learning objectives at this stage of my training
27. It is possible to do follow up with patients
28. There is enough time in the schedule for me to learn new skills
<b>Attendings' role</b>
29. My attendings take time to explain things when asked for advice
30. My attendings are happy to discuss patient care
31. There is (are) NO attending physician(s) who have a negative impact on the educational climate
32. My attendings treat me as an individual
33. My attendings treat me with respect
34. My attendings are all in their own way positive role models

35. When I need a attending, I can always contact one
36. When I need to consult an attending, they are readily available
<b>Formal education</b>
37. Residents are generally able to attend scheduled educational activities
38. Educational activities take place as scheduled
39. Attendings contribute actively to the delivery of high-quality formal education
40. Formal education and training activities are appropriate to my needs
<b>Role of the specialty tutor</b>
41. The specialty tutor monitors the progress of my training
42. The specialty tutor provides guidance to other attendings when needed
43. The specialty tutor is actively involved in improving the quality of education and training
44. In this rotation evaluations are useful discussions about my performance
45. My plans for the future are part of the discussion
46. During evaluations, input from several attendings is considered
<b>Patient sign-out</b>
47. When there is criticism of a management plan I have developed in consultation with my attending physician, I know the attending physician will back me up
48. Sign out takes place in a safe climate
49. Sign out is used as a teaching opportunity
50. Attendings encourage residents to join in the discussion during sign out

Figure 3 : première version du D-RECT par Boor et al.(28)

#### ➤ Deuxième version(8)

Compte tenu de l'importance croissante accordée au climat d'apprentissage et de la nécessité de nouvelles preuves de validité des instruments du climat d'apprentissage, il a été nécessaire de réévaluer les propriétés psychométriques initiales du D-RECT. Silkens and al.(8) ont retesté la validité interne et la fiabilité du D-RECT. Les résultats ont montré que le climat d'apprentissage pourrait être évalué à l'aide de 35 questions regroupées en neuf sous-échelles :



- Educational atmosphere
- Team work
- Role of speciality tutor
- Coaching and assessment
- Formal education
- Resident peer collaboration
- Work is adapted to residents competence
- Accessibility of supervisors
- Patient sign-out.

Cette nouvelle version a été administrée chez 1537 résidents aux Pays-Bas pour validation. Elle a montré une bonne consistance interne et un bon ajustement du modèle factoriel (répartition en neuf sous-domaines).

En outre, Silkens et al. ont montré que huit résidents par département ont été nécessaires pour évaluer toutes les sous-échelles du climat d'apprentissage de manière fiable.

Subscale	Item
<b>Educational atmosphere</b>	1. <i>Continuity of care is not affected by differences of opinion between attendings</i>
	2. <i>Differences of opinion between attendings about patient management are discussed in such a manner that is instructive to others present.</i>
	3. <i>Differences of opinion are not such that they have a negative impact on the work climate</i>
	4. <i>There is (are) NO attending physician(s) who have a negative impact on the educational climate</i>
	5. <i>My attendings treat me with respect.</i>
<b>Teamwork</b>	6. <i>Attendings, nursing staff, other allied health professionals and residents work together as a team.</i>
	7. <i>Nursing staff and other allied health professionals make a positive contribution to my training.</i>
	8. <i>Nursing staff and other allied health professionals are willing to reflect with me on the delivery of patient care</i>
<b>Role of specialty tutor</b>	9. <i>The specialty tutor monitors the progress of my training.</i>
	10. <i>The specialty tutor provides guidance to other attendings when needed</i>
	11. <i>The specialty tutor is actively involved in improving the quality of education and training.</i>
	12. <i>In this rotation evaluations are useful discussions about my performance.</i>
	13. <i>My plans for the future are part of the discussion.</i>
<b>Coaching and assessment</b>	14. <i>During evaluations, input from several attendings is considered.</i>
	15. <i>My attendings take the initiative to evaluate my performance.</i>
	16. <i>My attendings take the initiative to evaluate difficult situations I have been involved in.</i>
	17. <i>My attendings evaluate whether my performance in patient care is commensurate with my level of training</i>
	18. <i>My attendings occasionally observe me taking a history</i>
<b>Formal education</b>	19. <i>My attendings assess not only my medical expertise but also other skills such as teamwork, organization or professional behavior</i>
	20. <i>My attendings give regular feedback on my strengths and weaknesses</i>
	21. <i>Residents are generally able to attend scheduled educational activities</i>
	22. <i>Educational activities take place as scheduled</i>
<b>Resident peer collaboration</b>	23. <i>Attendings contribute actively to the delivery of high-quality formal education.</i>
	24. <i>Formal education and training activities are appropriate to my needs</i>
	25. <i>Résidents work well together</i>
<b>Work is adapted to residents' competence</b>	26. <i>Residents, as a group, make sure the day's work gets done.</i>
	27. <i>Within our group of residents, it is easy to find someone to cover or exchange a call</i>
	28. <i>The work I am doing is commensurate with my level of experience.</i>
<b>Accessibility of supervisors</b>	29. <i>The work I am doing suits my learning objectives at this stage of my training.</i>
	30. <i>It is possible to do follow up with patients</i>
	31. <i>When I need an attending, I can always contact one.</i>
<b>Patient sign-out</b>	32. <i>When I need to consult an attending, they are readily available.</i>
	33. <i>It is clear which attending supervises me.</i>
	34. <i>Sign-out is used as a teaching opportunity.</i>
	35. <i>Attendings encourage residents to join in the discussion during sign-out.</i>

Figure 4: Dutch Residency Education Climate Test (D-RECT). Version anglaise (8)

Globalement, la nouvelle structure correspond au questionnaire D-RECT initial. Bien que la sous-échelle «feedback» ait été supprimée, le sujet est toujours représenté dans le questionnaire par la sous-échelle «coaching and assessment». Les éléments de la sous-échelle «attending role» ont été divisés en nouvelles échelles «educational atmosphere» et «accessibility of supervisors». La littérature considère également que les enseignants-cliniciens sont principalement responsables de la création d'une atmosphère dans laquelle les apprenants peuvent facilement identifier et aborder leurs limites(31), ce qui est représenté dans le questionnaire par le facteur «educational atmosphere». La sous-échelle «Supervision» a été intégrée à la nouvelle sous-échelle «accessibility of supervisors». De même, l'accessibilité à une supervision opportune et appropriée a été considérée comme un facteur important dans les résultats à la fois en éducation et chez les patients(32).

En ce qui concerne le contenu des sous-échelles, les constructions du D-RECT s'inscrivent dans les cadres climatiques(33). Ostroff(34) a étudié le concept général du climat et a organisé les perceptions du climat en trois facettes d'ordre supérieur : la facette affective, la facette cognitive et la facette instrumentale.

Dans le D-RECT, la facette affective est représentée par le sentiment général de l'atmosphère (educational atmosphere), la qualité de la collaboration des résidents (resident peer collaboration) et la qualité de la collaboration de l'équipe (team work).

La facette cognitive est prise en compte par la manière dont le superviseur aide le résident à réfléchir sur sa performance (coaching and assessment), dans quelle mesure le résident participe à la passation du patient (patient sign-out) et dans quelle mesure le travail du résident est adapté au niveau d'expérience du résident (work is adapted to residents' competence).

Le D-RECT prend en compte la facette instrumentale en évaluant l'éducation planifiée (formal education), l'implication de l'éducateur officiel (role of the specialty tutor) et la mesure dans laquelle les superviseurs sont impliqués (accessibility of supervision). De plus, le D-RECT est un instrument opérationnel et basé sur la réalité du terrain. Le D-RECT étant un instrument permettant de mesurer le climat d'apprentissage de l'environnement post-gradué au travail, il est important que l'instrument ne se concentre pas uniquement sur les activités éducatives liées au résident et son travail, mais également sur les aspects liés aux soins des patients.

En résumé, l'instrument D-RECT a été développé en suivant une méthodologie rigoureuse et il a été validé sur un grand échantillon de résidents. Méthodologiquement, il s'agit de l'instrument le plus solide pour évaluer l'environnement de formation dans le cadre du résidanat.

#### ➤ **Impact de l'instrument D-RECT dans la littérature**

Plusieurs études ont utilisé le D-RECT pour évaluer l'environnement de formation et son impact sur la formation résidents. Ainsi le D-RECT a été utilisé aux Pays-Bas, Angleterre, Irlande, Australie, Allemagne, Pakistan, Colombie, Philippines et au Maroc(35-39).

Le D-RECT est utilisé aux Pays-Bas dans le cadre de la démarche qualité des programmes de résidanat(10). Cette démarche permet aux décideurs d'identifier spécifiquement de zones d'insuffisance et proposer des mesures correctives spécifiques. Dans ce cadre, Silkens et al. ont analysé en 2016 l'évolution de la perception de l'environnement de formation dans le temps aux Pays-Bas en utilisant le D-RECT. Les résultats ont montré une amélioration significative de la perception de l'environnement de formation par les résidents entre 2012 et 2014, suggérant que cette démarche pourrait avoir un impact positif sur la formation des résidents(10).

Van Vendeloo et al. ont étudié la relation entre la survenue du burn-out et la perception de l'environnement de formation en Belgique(40). L'étude a inclus 236 résidents de 29 spécialités en 2018. Les résultats ont montré une forte corrélation entre la survenue du Burn-out et la perception de l'environnement de formation par l'instrument D-RECT, suggérant que l'amélioration du climat de formation permettrait d'améliorer le bien-être des résidents.

Une étude a étudié la relation entre le climat de formation et l'efficacité de l'enseignement. Les résultats ont montré des associations positives entre le climat d'apprentissage global et les performances de l'enseignement. Plus spécifiquement, elle a montré des corrélations positives entre trois sous-domaines, à savoir le «Coaching and assessment», «Work adapted to residents competence» et «Formal education» avec l'efficacité de l'enseignement(41).

Dans le même sens, Silkens et al. ont analysé en 2018 la relation entre le climat de formation et les caractéristiques des services cliniques. Ce travail a montré que le type d'hôpital (Académique vs non-académique) et le pourcentage de temps consacré par l'enseignant aux activités pédagogiques prédisaient significativement les résultats du D-RECT(42). Ces résultats suggèrent que le D-RECT pourrait être utilisé pour identifier spécifiquement quelles actions il faudrait entreprendre pour améliorer la qualité de la prise en charge des résidents.

Pour ce qui est de la validation psychométrique du D-RECT en dehors des Pays-Bas, nous avons trouvé un seul travail dans lequel Pacifico et al. ont tenté de valider l'instrument dans le contexte du système de résidanat aux Philippines(37). Les auteurs n'ont pas pu valider le D-RECT dans sa forme originale (35 questions et 9 sous-domaines) à cause d'un ajustement du modèle inadéquat. Néanmoins, ils ont proposé une modification de l'instrument en éliminant 7 questions pour améliorer l'ajustement du modèle. Le résultat a été un questionnaire plus court de 28 questions réparties en 9 sous-domaines.

En synthèse, le D-RECT a montré un impact puissant dans l'analyse de l'environnement de formation dans la littérature. Il a été utilisé avec succès dans les démarches d'amélioration de la qualité des systèmes de résidanat. Il a également montré une forte corrélation avec la performance de l'enseignement dans les services cliniques et avec la survenue du burn-out chez les résidents. Cependant, la plupart de ces études ont été faites aux Pays-Bas et les propriétés psychométriques du D-RECT n'ont pas été validées dans d'autres contextes, surtout non européens. C'est pour ces raisons que nous avons entrepris ce travail de validation psychométrique de la traduction française du D-RECT, dans le contexte du résidanat au Maroc.

# MATÉRIEL ET MÉTHODES

## **A. Élaboration du questionnaire D-RECT en français**

### **1. Traduction**

Un comité fait de trois personnes a été constitué pour mener la traduction de l'instrument D-RECT, le thésard, et deux enseignants de médecine qui ont déjà publié sur le sujet de l'environnement de formation chez les résidents en médecine.

Le questionnaire original en anglais(8) a été traduit vers le français par un médecin bilingue en français et l'anglais. Ensuite, la version française a été traduite vers l'anglais par un traducteur professionnel. Les questionnaires traduits ont été évalués par le comité et une version initiale de l'instrument a été établie.

### **2. Prétest(43)**

Le prétest consiste en l'administration du questionnaire à un échantillon restreint de personnes susceptibles de faire partie de l'échantillon (médecins résidents) en vue d'assurer la clarté, la précision des termes utilisés. Le prétest est très important pour développer et affiner un instrument de mesure(44) et est considéré comme étant l'un des aspects importants de la qualité de la recherche par enquête(45). Il a plusieurs objectifs : il permet de mettre à l'épreuve la forme des questions et leur ordonnancement, de vérifier la compréhension des répondants, d'examiner la pertinence des modalités de réponses proposées.

La version 1 a ainsi été soumise à 10 résidents. Pour chaque question, les résidents devaient indiquer si l'intitulé était clair et bien compris et proposer des modifications pour améliorer le questionnaire. Toutes les remarques ont été notées puis évaluées par le comité de coordination pour aboutir à la version définitive du questionnaire en français (Tableau 5).



**Tableau 5 : Version traduite en français du D-RECT**

Sous-échelle	Élément
Atmosphère éducative	1. La continuité des soins des patients n'est pas affectée par les différences d'opinions entre les encadrants (enseignants)
	2. Les différences d'opinions entre les encadrants (enseignants) au sujet de la prise en charge des patients sont discutées de manière instructive pour les autres personnes présentes.
	3. Les différences d'opinions n'ont pas un impact négatif sur le climat de travail
	4. Aucun encadrant n'a un impact négatif sur le climat de formation.
	5. Mes encadrants me traitent avec respect.
Le travail en équipe	6. Les encadrants, le personnel infirmier, les autres professionnels de la santé et les résidents travaillent ensemble en équipe.
	7. Le personnel infirmier et les autres professionnels de la santé apportent une contribution positive à ma formation.
	8. Le personnel infirmier et les autres professionnels de la santé sont disposés à réfléchir avec moi sur la prestation des soins aux patients
Rôle du chef de service	9. le chef de service surveille les progrès de ma formation.
	10. Le chef de service fournit des conseils à d'autres encadrants en cas de besoin.
	11. Le chef de service est activement impliqué dans l'amélioration de la qualité de l'enseignement et de la formation.
	12. Durant ce stage, les évaluations incluent des discussions utiles sur mes performances.
	13. Mes projets pour l'avenir font partie de ces discussions (lors des évaluations).
Coaching et évaluation	14. L'avis de plusieurs encadrants est pris en considération lors des évaluations
	15. Mes encadrants prennent l'initiative d'évaluer ma performance.
	16. Mes encadrants prennent l'initiative d'évaluer les situations difficiles dans lesquelles j'ai été impliqué.
	17. Mes encadrants évaluent si ma performance dans la prise en charge des patients correspond à mon niveau de formation
	18. Mes encadrants m'observent occasionnellement quand j'interroge un patient

	19. Mes encadrants évaluent non seulement mon expertise médicale, mais aussi d'autres compétences telles que le travail d'équipe, l'organisation ou le comportement professionnel
	20. Mes encadrants donnent des Feedbacks réguliers sur mes forces et mes faiblesses
<b>Enseignement et formation</b>	21. Les résidents sont généralement en mesure d'assister aux activités d'enseignement programmées (exemple : cours, staffs...)
	22. Les activités d'enseignements se déroulent comme planifié.
	23. Les encadrants contribuent activement à la présentation d'un contenu académique de grande qualité.
	24. Les activités d'enseignement et de formation académiques sont appropriées à mes besoins
<b>Collaboration entre résidents</b>	25. Les résidents travaillent bien ensemble
	26. Les résidents, en tant que groupe, s'assurent que le travail de la journée a été réalisé.
	27. Au sein de notre groupe de résidents, il est facile de trouver quelqu'un pour assurer ou échanger une garde
<b>Le travail est adapté à la compétence des résidents.</b>	28. Le travail que je fais est adapté à mon niveau d'expérience.
	29. Le travail que je fais correspond à mes objectifs d'apprentissage à ce stade de ma formation.
	30. Il est possible de suivre l'évolution des patients.
<b>Accessibilité des encadrants</b>	31. Lorsque j'ai besoin d'un encadrant, je peux toujours en contacter un.
	32. Quand j'ai besoin d'avoir l'avis d'un encadrant, ils sont facilement disponibles.
	33. L'encadrant qui me supervise est clairement identifié
<b>La sortie du patient</b>	34. La sortie des patients est utilisée comme une opportunité d'enseignement
	35. Les encadrants encouragent les résidents à participer à la discussion lors de la sortie des patients

## **B. Distribution du questionnaire**

Le questionnaire final comporte, en plus de la version française du D-RECT, les questions démographiques et professionnelles suivantes :

- Sexe (homme/femme).
- Âge en année.
- Ville.
- Année de résidanat.
- Accès au résidanat (Internat, Concours de résidanat, Agence marocaine de Coopération internationale).
- Statut (bénévole, contractuel, attaché au CHU, étranger).
- Type de spécialité (médicale, chirurgicale, médico-chirurgicale).
- Spécialité.

Le questionnaire informatisé a été établi en utilisant la plateforme Google Form (<https://www.google.com/forms/about/>). L'utilisation des questionnaires informatisés est plus avantageuse, car elle permet de réduire les erreurs de saisie et d'accroître le taux de réponse(46).

La version définitive du questionnaire se trouve sur le lien suivant :<https://goo.gl/forms/ZPR78aoJjpw4fOIQ2>.

## **C. Administration du questionnaire**

Le questionnaire en ligne a été soumis aux résidents marocains entre le 1er juillet et le 30 septembre 2018. L'objectif initial a été de proposer le questionnaire aux résidents des centres hospitaliers universitaires (CHU) de Fès, Oujda, Rabat, Casablanca et Marrakech.

En l'absence d'une base de données de courriers électroniques ou postaux des résidents dans les facultés de médecine, il n'a pas été possible de cibler directement des résidents de façon homogène. Pour remédier à cette difficulté, des résidents-référents ont été désignés dans chaque CHU et ils se sont chargés de diffuser le lien vers la page web du formulaire auprès des résidents.

La participation à l'étude était volontaire et les données ont été recueillies anonymement.

## **D. Méthodes statistiques**

### **1. Statistiques descriptives**

Les variables quantitatives sont exprimées en moyennes et déviations standards, ou en médianes et quartiles, selon la normalité de la distribution de la population. La normalité de distribution des données a été évaluée par le test d'asymétrie et le test de Kurtosis. Des valeurs absolues d'asymétrie inférieure à 3 et de Kurtosis inférieur à 10 sont considérées comme acceptables pour une analyse factorielle confirmatoire(47).

Les variables quantitatives sont exprimées en nombre et pourcentage.

Pour les résultats du score D-RECT, la moyenne générale avec déviation standard de toutes les questions, et les moyennes et déviations standards de chacun des neuf sous-domaines sont présentées.

Les données manquantes ont été testées pour examiner le caractère aléatoire ou systématique du manque de données. Ceci a été fait par le test de Khi-deux de Little MCAR (Missing Completely at Random). Ensuite, ces valeurs manquantes ont été remplacées en utilisant la technique d'expectation-maximisation (EM).

## **2. Analyse de la fidélité(43,47)**

La fidélité d'un test concerne la précision de l'instrument, peu importe ce qu'il mesure. Un test fidèle mesure toujours le construit psychologique de la même manière. Plusieurs indices de fidélité ont été proposés dans la littérature. Les deux plus fréquents sont la consistance interne et la stabilité temporelle (ou test—retest).

### **a. Consistance interne(48)**

La consistance interne concerne la capacité d'un instrument à mesurer de manière cohérente. La consistance interne décrit la mesure dans laquelle tous les éléments d'un test mesurent le même concept ou la même construction et sont donc liés à la corrélation des éléments dans le test. De plus, les estimations de consistance montrent la quantité d'erreurs de mesure dans un test. Il convient de noter que la consistance d'un instrument est étroitement liée à sa validité(48,49). Un instrument ne peut pas être valide sauf s'il est fiable.

Le test de consistance interne est le plus fréquemment mesuré par le test Alpha de Cronbach. Le résultat est exprimé en chiffre entre 0 et 1. Un résultat supérieur à 0,7 est considéré comme satisfaisant(50).

Pour ce travail, la consistance interne est mesurée pour le test D-RECT entier et pour chacun des neuf sous-domaines.

### **b. Test —Retest(43)**

La deuxième mesure de fidélité porte sur les propriétés de stabilité temporelle de l'instrument. Elle se mesure en demandant à des sujets de répondre à l'instrument à deux reprises avec un écart de temps relativement important entre les deux réponses. Si l'instrument est fidèle, les réponses des sujets devraient demeurer stables entre les deux passations de l'instrument, à condition qu'ils n'aient pas changé de contexte (service). Des indices de corrélation test-retest supérieurs à 0,6 sont satisfaisants.

Nous avons demandé à 15 résidents de répondre au questionnaire une deuxième fois afin d'évaluer la stabilité des réponses dans le temps. Un minimum de deux semaines devait d'écouler entre la première et la deuxième fois qu'ils ont répondu. Nous avons utilisé le coefficient de corrélation intra-classe avec un modèle mixte à deux facteurs de type cohérence absolue.

### **3. Evaluation de la validité du construit(43)**

L'évaluation de la validité du construit de la version traduite consiste à vérifier si celle-ci permet de bien mesurer le construit, tel que défini par son cadre théorique spécifique. Cette évaluation vise aussi à s'assurer que la version traduite est assez sensible pour déceler les effets du construit hypothétique propre à la théorie. Cette évaluation est importante, car il est possible qu'un questionnaire possède une validité concomitante et de contenu, sans pour autant manifester une validité du construit. Il est important de démontrer que la théorie qui sous-tend l'instrument est valide dans l'environnement du système de résidanat au Maroc, sinon l'utilité du questionnaire et la validité de la théorie sous-jacente à ce dernier seraient remise en question.

### **a. Structure du construit : analyse factorielle confirmatoire(43,51,52)**

L'évaluation de la validité du construit d'un questionnaire consiste à démontrer que ce dernier est bel et bien formulé de façon conséquente à la théorie qui le sous-tend. Dans ce contexte l'analyse factorielle peut s'avérer pertinente pour tester la validité de construit de la version traduite en français. En effet, dans la mesure où ce type d'analyse permet d'étudier les regroupements entre les items, il devient possible de vérifier si les résultats démontrent bel et bien la présence de 9 facteurs correspondant aux sous-domaines proposés par la version originale du questionnaire(8).

Il existe deux types d'analyses factorielles utilisées dans la littérature, l'analyse factorielle exploratoire (AFE) et l'analyse factorielle confirmatoire (AFC). Cette dernière est actuellement la méthode statistique la plus utilisée pour valider la structure d'un questionnaire(53). Une AFC consiste à confronter les données observées à un modèle théorique de façon à déterminer dans quelle mesure le modèle théorique correspond aux données recueillies. Une structure factorielle et donc définie a priori, puis son degré d'adéquation aux données sont calculés. Dans notre cas, nous avons confronté les données recueillies chez les résidents marocains au modèle théorique à neuf facteurs (ou sous-domaines) proposé par des auteurs du questionnaire original(8).

#### **➤ Représentation du modèle AFC (figure AFC)**

Dans la représentation graphique d'un modèle AFC, les variables observées (questions du questionnaire) sont représentées par des rectangles. Les variables latentes sont les variables non observées représentées par les construits (ou sous-domaines du questionnaire). Elles sont représentées par des ovales.

Les paramètres du modèle sont les caractéristiques de la population qui vont être estimées et testées par l'AFC. Les relations entre les variables observées et latentes sont indiquées par des flèches allant des variables latentes vers les variables observées. Cette direction des flèches indique la prédiction que les construits (ou sous-domaine) sous-jacent causent les variables observées (réponses au questionnaire). Les facteurs générés sont des coefficients de régression pour la prédiction des indicateurs à partir du facteur latent. En général, quand le facteur généré est supérieur à 0,71, il est considéré excellent, supérieur à 0,63 est considéré très bon, 0,55 est considéré bon, supérieur à 0,45 est considéré acceptable(54).

Les flèches entre les variables latentes, ou facteurs représentent les corrélations entre les facteurs. Brown(51) a proposé que les corrélations entre construits de 0,85 et plus indiquent une faible validité discriminatoire.

➤ **Indicateurs d'évaluation de la structure du construit**

Il existe plusieurs indicateurs pour évaluer la structure du construit. Brown(51) identifie quatre catégories d'indices : indices absolus, indices de parcimonie, indices de comparaisons et indices de prédiction.



### ➤ **Indices absolus**

Les indices absolus testent l'hypothèse que la matrice variance-covariance prédite est égale à la matrice de variance-covariance de l'échantillon. Trois indicateurs absolus sont décrits :

- Le Khi-deux corrigé : est indicateur absolu le plus utilisé. Il teste si le modèle est adapté à la population. Cet indicateur a plusieurs limitations notamment qu'il dépend de la taille de l'échantillon et qu'il sera systématiquement significatif avec des échantillons larges.
- RMR (Root Mean Square Residual) : disparité moyenne entre les covariances de l'échantillon et les covariances prédites par le modèle théorique. Il est difficile à interpréter.
- SRMR (Standardized Root Mean Square Residual) : il est basé sur la disparité des corrélations dans la matrice de l'échantillon et les corrélations prédites par le modèle. C'est un indice standardisé et généralement préféré au RMR.

### ➤ **Indices de parcimonie**

- RMSEA (Root Mean Square Error of Approximation) : il évalue à quel point le modèle s'adapte raisonnablement à la population. Il est sensible à la complexité du modèle et insensible à la taille de l'échantillon.

### ➤ **Indices de comparaison**

Ils sont utilisés pour comparer l'adéquation entre plusieurs modèles. Des exemples sont :

- CFI : Comparative Fit Index.
- TLI : Tucker-Lewis Index.
- NNFI: Non-Normed Fit Index.

### ➤ Indices de prédiction

Ces indices permettent d'évaluer le modèle dans des échantillons de réplification hypothétique de même taille et tiré au hasard dans la même population que l'échantillon original. Ces indices pourraient être vus comme basés sur la population plutôt que sur l'échantillon(47).

- *AIC (Akaike Information Criterion) : est généralement utilisé pour comparer entre deux (ou plus) modèles sur le même set de données. Un AIC bas suggère que le modèle est adéquat. Ainsi, quand deux modèles sont comparés, celui avec l'AIC le plus bas est choisi comme meilleur modèle.*
- *ECVI (Expected Cross-validation Index) : est aussi utilisé pour comparer les modèles et est interprété de la même manière que l'AIC.*

Les valeurs seuils pour chaque indice sont présentées dans le tableau suivant(51,52) :

**Tableau des Indices d'ajustement pour analyse factorielle confirmatoire**

Indice	Valeurs critiques
<b>Indices absolus</b>	
Khi-deux corrigé	$P \geq 0.05$
Khi-deux normé	$\leq 2$
SRMR	< 0,08 bon < 0,12 acceptable
<b>Indices de parcimonie</b>	
RMSEA	$\leq 0,50$ bon $\leq 0.80$ raisonnable
<b>Indices de comparaison</b>	
CFI	> 0,90 acceptable > 0,95 bon
TLI	> 0,90 acceptable > 0,95 bon
NNFI	> 0,90
<b>Indices de prédiction</b>	
AIC	Bas
ECVI	Bas
GFI	$\geq 0.90$
AGFI	$\geq 0.80$

### **b. Validité convergente**

La validité convergente représente le degré de convergence entre de multiples tentatives de mesurer le même concept (ou sous-domaine) avec des méthodes différentes(51). Elle correspond au rapport entre la variance des indicateurs de mesure expliquée par le concept latent et la variance totale incluant les erreurs de mesure (erreur aléatoire et erreur systématique). On appelle ce ratio Average Variance Extracted (AVE). S'il est inférieur à 50 %, cela signifie que la variance due aux erreurs de mesure est supérieure à celle expliquée par le concept lui-même.

### **c. Validité discriminatoire**

La validité discriminante est destinée à s'assurer que les indicateurs de mesure d'un construit sont faiblement corrélés aux indicateurs de mesure d'autres construits. La validité discriminatoire vise à évaluer la corrélation entre les mesures de différents concepts ou (sous-domaines). Brown(51) a proposé que les corrélations entre construits de 0,85 et plus indiquaient une faible validité discriminatoire.

## **4. Logiciel statistique**

L'application IBM SPSS Statistics version 21 a été utilisée pour les statistiques descriptives, l'analyse de la consistance interne et le test—retest.

L'application SPSS Amos version 21 a été utilisée pour l'analyse factorielle confirmatoire.

# RÉSULTATS

Durant la période d'études, sur 695 résidents contactés, 211 ont répondu au questionnaire ce qui fait un taux de réponse de 30,35 %. Durant la première semaine de l'étude, beaucoup de résidents ont exprimé une réticence par rapport à la question « Spécialité » avec l'argument qu'elle permettrait d'identifier les résidents et ainsi violer l'anonymat. Après concertation, le comité de pilotage a décidé d'enlever cette question du formulaire en ligne.

## **A. Description de la population**

Les caractéristiques des participants sont présentées dans le tableau 2.

### **1. Profil démographique**

L'âge moyen des participants a été 29,11 ans (minimum 23, maximum 40).

Il y avait homme et femme

Les résidents des CHU de Rabat (44,5 %) et de Fès (40,3 %) ont été les plus représentés dans ce travail.

### **2. Profil professionnel**

Les résidents de première année (25,6 %), troisième année (24,6 %) et quatrième année (23,7 %) étaient les plus représentés dans l'étude.

Il y avait légèrement plus de résidents admis suite au concours de résidanat (51,2 %) que de résidents admis par la voie de l'internat du CHU (42,7 %).

Quatorze résidents étrangers ont participé à l'étude (6,6 %) dont 13 via l'agence marocaine de coopération internationale.

Concernant le type de contrat des résidents, il y avait plus de résidents contractuels (contrat avec le ministère de la Santé) que de résidents bénévoles (contrat limité à la durée de la formation) (48,3 % vs 40,8 %).

Finalement, il y avait plus résidents en spécialité médicale (54,5 %) qu'en spécialité chirurgicale (32,2 %) ou médico-chirurgicale (13,3 %).

**Tableau 2 : Caractéristiques démographiques de la population**

Variables	Résultats
<b>Âge moyen</b> (années, écart-type)	29.11 (2,65)
<b>Sexe</b>	
Hommes	97 (46,0 %)
Femmes	114 (54,0 %)
<b>Ville</b>	
Fès	85 (40,3 %)
Rabat	94 (44,5 %)
Marrakech	8 (3,8 %)
Casablanca	13 (6,2 %)
Oujda	11 (5,2 %)
<b>Année de résidanat</b>	
1ère année	54 (25,6 %)
2ème année	34 (16,1 %)
3ème année	52 (24,6 %)
4ème année	50 (23,7 %)
5ème année	21 (10,0 %)
<b>Accès au résidanat</b>	
Concours	108 (51,2 %)
Titre	90 (42,7 %)
Étrangers	13 (6,2 %)
<b>Type de contrat</b>	
Contractuel	86 (40,8 %)
Bénévole	102 (48,3 %)
Attaché au CHU	9 (4,3 %)
Étrangers	14 (6,6 %)
<b>Type de spécialité</b>	
Médicale	115 (54,5 %)
Chirurgicale	68 (32,2 %)
Médico-chirurgicale	28 (13,3 %)

## B. Résultats des scores D-RECT

Sur les 211 réponses reçues, il y avait un formulaire qui contenait deux réponses manquantes chez le même résident, ce qui nous fait moins de 0,01 % de données manquantes.

Le score moyen du score D-RECT a été 3,21 (déviations standard 0.77). Le tableau 3 montre les résultats des 9 sous-domaines du D-RECT.

**Tableau 3 : Résultats des sous-domaines du D-RECT.**

Sous-domaine	Score moyen	Déviations standard
<b>Atmosphère éducative</b>	3,16	0,92
<b>Travail d'équipe</b>	3,07	0,99
<b>Rôle du chef de service</b>	3,21	1,02
<b>Coaching et évaluation</b>	3,00	1,05
<b>Enseignement</b>	3,10	1,04
<b>Collaboration entre résidents</b>	3,54	0,90
<b>Travail adapté aux compétences des résidents</b>	3,19	0,90
<b>Disponibilité des encadrants</b>	3,79	0,95
<b>Sortie des patients</b>	3,04	1,14

Les scores moyens par question ont varié de 2,66 (Question 4 atmosphère) à 3.94 (Question encadrant 1). Le tableau 4 montre les scores moyens de chaque question du D-RECT. Les cinq scores les plus bas (rouge) et les plus hauts (vert) sont mis en gras dans le tableau.



**Tableau 4 : Scores moyens des questions du questionnaire D-RECT**

Question	Moyenne	Écart-type
Atmosphère éducative 1	3 213	1,1780
Atmosphère éducative 2	3 531	1,1640
Atmosphère éducative 3	2 919	1,2529
Atmosphère éducative 4	<b>2 659</b>	1,2899
Atmosphère éducative 5	3 493	1,1396
Travail d'équipe 1	3 185	1,1335
Travail d'équipe 2	3 114	1,1025
Travail d'équipe 3	2 924	1,2961
Rôle du chef de service 1	3 289	1,2861
Rôle du chef de service 2	<b>3 562</b>	1,1163
Rôle du chef de service 3	3 427	1,3159
Rôle du chef de service 4	3 085	1,2315
Rôle du chef de service 5	<b>2 682</b>	1,2026
Rôle du chef de service 6	3 199	1,1581
Coaching et évaluation 1	3 180	1,1775
Coaching et évaluation 2	3 047	1,2022
Coaching et évaluation 3	3 180	1,2094
Coaching et évaluation 4	<b>2 716</b>	1,2477
Coaching et évaluation 5	3 071	1,3272
Coaching et évaluation 6	<b>2 847</b>	1,3222
Enseignement 1	3 531	1,1224
Enseignement 2	<b>2 877</b>	1,2047
Enseignement 3	3 052	1,2657

Enseignement 4	2 938	1,2193
Collaboration entre résidents 1	3 531	1,0703
Collaboration entre résidents 2	3 526	1,1011
Collaboration entre résidents 3	<b>3 573</b>	1,0038
Travail adapté aux compétences des résidents 1	2 924	1,1271
Travail adapté aux compétences des résidents 2	2 957	1,2318
Travail adapté aux compétences des résidents 3	<b>3 678</b>	1,0422
Disponibilité des encadrants 1	<b>3 943</b>	.9693
Disponibilité des encadrants 2	<b>3 787</b>	1,0314
Disponibilité des encadrants 3	<b>3 668</b>	1,2776
Sortie des patients 1	2 957	1,1764
Sortie des patients 2	3 123	1,2436

#### Tests de normalité (tableau 5)

Les tests de normalité ont montré des valeurs absolues d'asymétrie inférieure à 3 et de Kurtosis inférieure à 10 et sont donc considérées comme acceptables pour une analyse factorielle confirmatoire

**Tableau 5 : Analyse de la distribution par les tests d'asymétrie et de Kurtosis.**

	Asymétrie		Kurtosis	
	Statistiques	Erreur std.	Statistiques	Erreur std.
Atmosphère éducative 1	-.369	.167	-.895	.333
Atmosphère éducative 2	-.807	.167	-.161	.333
Atmosphère éducative 3	-.081	.167	-1 095	.333
Atmosphère éducative 4	.108	.167	-1 280	.333
Atmosphère éducative 5	-.625	.167	-.378	.333
Travail d'équipe 1	-.350	.167	-.764	.333
Travail d'équipe 2	-.464	.167	-.666	.333
Travail d'équipe 3	-.216	.167	-1 272	.333
Rôle du chef de service 1	-.446	.167	-.920	.333
Rôle du chef de service 2	-.748	.167	-.210	.333
Rôle du chef de service 3	-.548	.167	-.871	.333
Rôle du chef de service 4	-.303	.167	-1 002	.333
Rôle du chef de service 5	.103	.167	-.977	.333
Rôle du chef de service 6	-.321	.167	-.773	.333
Coaching et évaluation 1	-.390	.167	-.901	.333
Coaching et évaluation 2	-.142	.167	-1 051	.333
Coaching et évaluation 3	-.302	.167	-.931	.333
Coaching et évaluation 4	.004	.167	-1 285	.333
Coaching et évaluation 5	-.304	.167	-1 120	.333
Coaching et évaluation 6	.060	.167	-1 231	.333
Enseignement 1	-.873	.167	-.013	.333
Enseignement 2	-.124	.167	-1 103	.333

Enseignement 3		-.227	.167	-1 066	.333
Enseignement 4		-.152	.167	-1 040	.333
Collaboration résidents 1	entre	-.669	.167	-.180	.333
Collaboration résidents 2	entre	-.747	.167	-.127	.333
Collaboration résidents 3	entre	-.688	.167	.049	.333
Travail adapté compétences résidents 1	aux des	-.273	.167	-.912	.333
Travail adapté compétences résidents 2	aux des	-0,196	0,167	-1 020	0,333
Travail adapté compétences résidents 3	aux des	-0,979	0,167	0,598	0,333
Disponibilité encadrant1	des	-0,994	0,167	0,660	0,333
Disponibilité encadrant2	des	-0,719	0,167	-0,101	0,333
Disponibilité des 3		-0,906	0,167	-0,220	0,333
Sortie du patient 1		-0,112	0,167	-0,819	0,333
Sortie du patient 2		-0,266	0,167	-0,948	0,333

## C. Analyse de la fidélité

### 1. Consistance interne

Le tableau 6 montre le résultat du test Cronbach alpha pour le score D-RECT et pour chacune des 9 dimensions du score. Toutes les dimensions ont montré un score satisfaisant supérieur à 0,7.

**Tableau 6 : Test de Cronbach alpha pour les dimensions du score D-RECT.**

Dimension	Score Cronbach
D-RECT	0,95
Atmosphère éducative	0,82
Travail d'équipe	0,79
Rôle du chef de service	0,91
Coaching et évaluation	0,91
Enseignement	0,89
Collaboration entre résidents	0,81
Travail adapté aux compétences des résidents	0,70
Disponibilité des encadrants	0,82
Sortie du patient	0,88

### 2. Test-retest

Sur les quinze résidents contactés, treize ont répondu deux fois au questionnaire. Le coefficient de corrélation obtenu a été 0,89, indiquant une bonne corrélation entre les deux mesures.

## **D. Analyse factorielle confirmatoire**

Le diagramme 2 montre le diagramme généré par l'application SPSS Amos pour l'analyse factorielle confirmatoire.

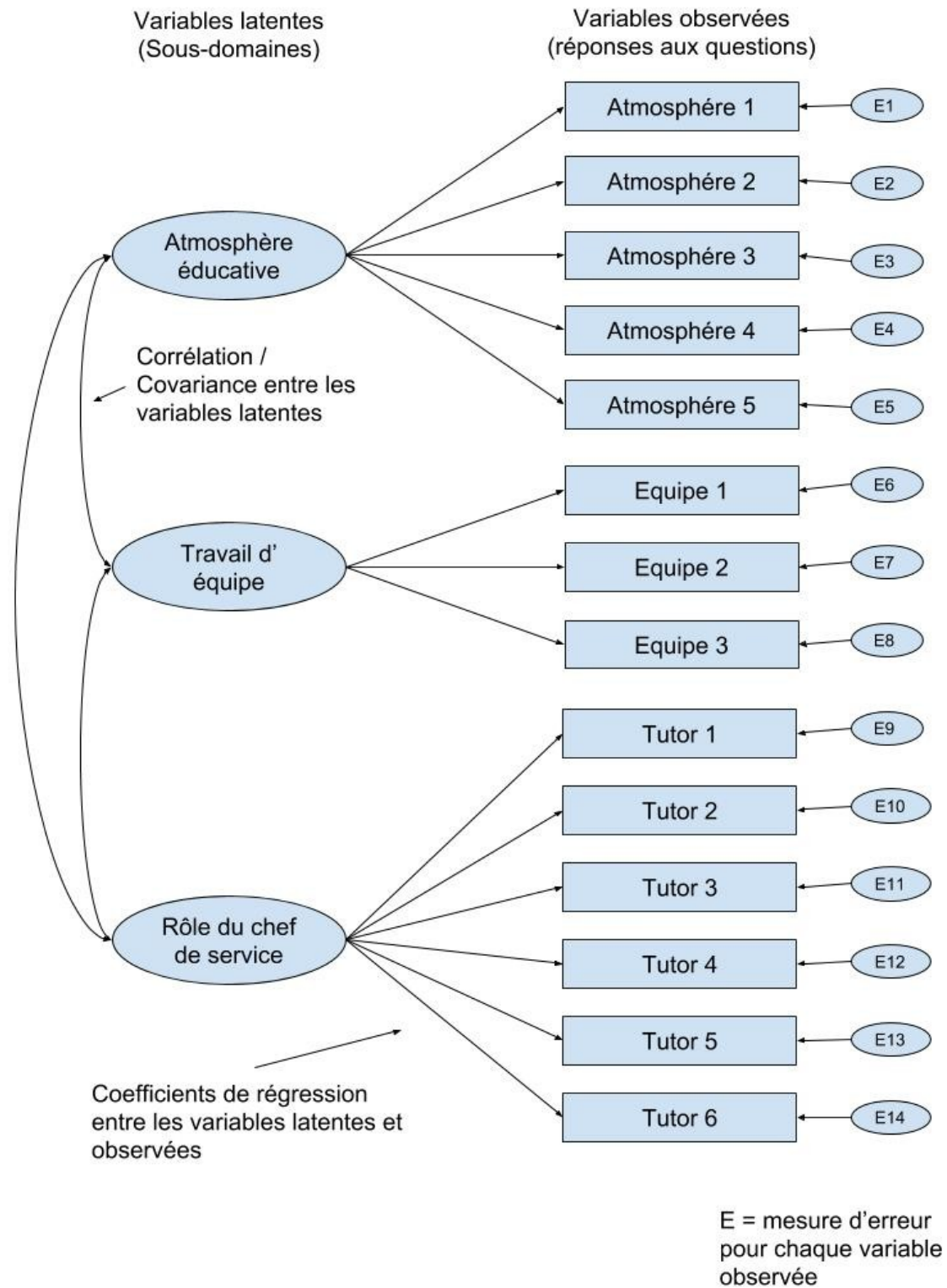


Figure 5 : Représentation partielle du diagramme de l'analyse factorielle confirmatoire

## 1. Validité du construit

Les coefficients de régression (entre les variables observées et les variables latentes) sont montrés sur le tableau 7. Tous les coefficients de régression sont considérés comme bons, à l'exception du coefficient de la question 3 de la dimension « Travail adapté aux résidents » qui a un score acceptable (0,47).

**Tableau 7 : Coefficients de régression de l'analyse factorielle confirmatoire.**

			Coefficients de régression
Atmosphère3	<---	atmosphere	0,71
Atmosphère2	<---	atmosphere	0,68
Atmosphère1	<---	atmosphere	0,63
Atmosphère4	<---	atmosphere	0,73
Atmosphère5	<---	atmosphere	0,71
teamwork3	<---	teamwork	0,78
teamwork2	<---	teamwork	0,81
teamwork1	<---	teamwork	0,69
Tutor6	<---	tutor	0,67
Tutor5	<---	tutor	0,76
Tutor4	<---	tutor	0,84
Tutor3	<---	tutor	0,86
Tutor2	<---	tutor	0,79
Tutor1	<---	tutor	0,86
Coaching6	<---	coaching	0,81
Coaching5	<---	coaching	0,79
Coaching4	<---	coaching	0,67
Coaching3	<---	coaching	0,84
Coaching2	<---	coaching	0,86
Coaching1	<---	coaching	0,84
Enseignement4	<---	teaching	0,89
Enseignement3	<---	teaching	0,88
Enseignement2	<---	teaching	0,76
Enseignement1	<---	teaching	0,72
Collaboration3	<---	collaboration	0,65
Collaboration2	<---	collaboration	0,79



Collaboration1	<---	collaboration	0,86
Adapt3	<---	adapt	0,47
Adapt2	<---	adapt	0,82
Adapt1	<---	adapt	0,76
Encadrant3	<---	encadrant	0,63
Encadrant2	<---	encadrant	0,91
Encadrant1	<---	encadrant	0,88
Sortie2	<---	sortie	0,92
Sortie1	<---	sortie	0,85

Le **tableau 8** montre les indicateurs d'évaluation de la structure du construit. En se basant sur les recommandations de Brown, nous avons deux indicateurs qui ont des résultats satisfaisants (SRMR et RMSEA) et deux indicateurs qui ont un résultat légèrement sous le seuil recommandé (CFI et TLI).

**Tableau 8 : indicateurs de la validité du construit de l'analyse factorielle confirmatoire**

Indice	Résultat	Valeurs critiques
<b>Indices absolus</b>		
Khi-deux corrigé	< 0,05	$P \geq 0.05$
Khi-deux normé	2,15	$\leq 2$
<b>SRMR</b>	<b>0,058</b>	< 0,08 bon < 0,12 acceptable
<b>Indices de parcimonie</b>		
<b>RMSEA</b>	<b>0,07</b>	$\leq 0,50$ bon $\leq 0.80$ raisonnable
<b>Indices de comparaison</b>		
<b>CFI</b>	<b>0,88</b>	> 0,90 acceptable > 0,95 bon
<b>TLI</b>	<b>0,87</b>	> 0,90 acceptable > 0,95 bon
NNFI	0,80	> 0,90
<b>Indices de prédiction</b>		
AIC	1328,22	Bas
ECVI	6,37	Bas
GFI	0,76	$\geq 0,90$
AGFI	0,71	$\geq 0,80$

## 2. Validité convergente

Toutes les dimensions de l'instrument ont un score AVE supérieur à 0,5, à l'exception de la dimension « Travail d'équipe » qui a score légèrement inférieur (0,48).

**Tableau 9 : Étude de la validité convergente**

Dimensions du D-RECT	Average variance extracted
Atmosphère éducative	0,67
Travail d'équipe	0,48
Rôle du chef de service	0,58
Coaching et évaluation	0,65
Enseignement	0,65
Collaboration entre résidents	0,67
Travail adapté aux compétences des résidents	0,60
Disponibilité des encadrants	0,50
Sortie du patient	0,79

## 3. Validité discriminatoire

L'analyse des corrélations entre les dimensions du D-RECT a donné des facteurs de corrélation inférieurs à 0,85, témoignant d'une bonne validité discriminatoire.

## **E. Ajustement du modèle**

Vu que certains indicateurs de validité du construit étaient inférieurs aux recommandations de Brown, nous avons essayé d'ajuster le modèle du construit en se basant toujours sur les recommandations de Brown.

Nous avons analysé les indices de modifications pour chercher ceux qui seraient élevés. Ensuite, nous avons ajouté des liens de covariance entre les variables observées qui montraient des indices de modifications élevés, appartenant à la même dimension (ou facteur) :

- Atmosphère éducative : entre les questions 1 et 3.
- Rôle du chef de service : entre les questions 1 et 2.
- Rôle du chef de service : entre les questions 1 et 3.
- Rôle du chef de service : entre les questions 2 et 3.
- Coaching et évaluation : entre les questions 2 et 3.

Le tableau 10 montre les résultats des indicateurs de validité du construit du nouveau modèle, comparé au modèle original. Nous remarquons que, en gardant la même structure du construit et les mêmes variables observées, nous obtenons une meilleure adéquation du modèle sur les indices d'ajustement. Néanmoins, la différence entre les deux modèles n'est pas significative ( $p > 0,05$ ).

**Tableau 10 : Indicateurs de validité du construit du modèle original et du modèle modifié**

Indice	Modèle modifié	Modèle original	Valeurs critiques
<b>Indices absolus</b>			
Khi-deux corrigé	< 0,05	< 0,05	$P \geq 0.05$
Khi-deux normé	1,93	2,15	$\leq 2$
SRMR	<b>0,055</b>	<b>0,058</b>	< 0,08 bon < 0,12 acceptable
<b>Indices de parcimonie</b>			
RMSEA	<b>0,06</b>	<b>0,07</b>	$\leq 0,50$ bon $\leq 0.80$ raisonnable
<b>Indices de comparaison</b>			
CFI	<b>0,91</b>	<b>0,88</b>	> 0,90 acceptable > 0,95 bon
TLI	<b>0,89</b>	<b>0,87</b>	> 0,90 acceptable > 0,95 bon
NNFI	0,82	0,80	> 0,90
<b>Indices de prédiction</b>			
AIC	1227,19	1328,22	Bas
ECVI	5,84	6,37	Bas
GFI	0,78	0,76	$\geq 0,90$
AGFI	0,74	0,71	$\geq 0,80$

# DISCUSSION

Bien que l'environnement de formation est un concept qui a gagné en popularité dans le monde anglo-saxon, son utilisation dans le monde francophone reste limitée, essentiellement à cause de l'absence de validation d'instruments en langue française. Nous avons donc essayé de répondre à cette problématique à travers ce travail en validant la version française du Dutch Résidency Educational Climate Test (D-RECT) dans le contexte spécifique du résidanat au Maroc.

Après traduction de l'instrument D-RECT en français, nous avons pu obtenir 211 réponses au questionnaire. L'analyse statistique a montré que notre version traduite avait une bonne consistance interne, une bonne fiabilité temporelle. L'analyse factorielle confirmatoire a permis de démontrer une validité du construit acceptable avec une bonne validité convergente et discriminatoire.

Nous allons discuter chacune des étapes de validation du questionnaire en confrontant notre méthodologie par rapport à ce qui se fait dans la littérature. Ensuite, nous discuterons concrètement les utilisations possibles du questionnaire D-RECT, essentiellement les possibilités de recherche futures.

## **A. Méthodologie de validation**

La validation psychométrique d'une traduction d'un instrument est un processus rigoureux qui nécessite plusieurs étapes, encadrées par des recommandations internationales. Notre validation est basée sur les recommandations de Vallerand(43), qui par l'intermédiaire de ses recherches en est venu à proposer une méthodologie de validation transculturelle des instruments de nature psychométrique. Cette méthodologie permet aux chercheurs de langue française de s'approprier des instruments de langue anglaise afin de s'en servir pour des fins de recherche.

## 1. Traduction

La traduction est l'une des premières étapes du processus de validation. Elle nécessite la préparation d'une version préliminaire du questionnaire original dans la langue cible. De nombreuses méthodes peuvent être utilisées pour obtenir une version préliminaire du questionnaire adéquate. Selon l'avis de plusieurs chercheurs en psychologie transculturelle(43,55), la technique de la traduction inversée (back-translation) semble idéale pour établir une version préliminaire dans la langue cible.

Les avantages de la traduction inversée sont au moins de deux ordres :

- Premièrement, ce type de traduction implique la participation de deux personnes. Il y a donc moins de chances que les partis pris, les préjugés, les interprétations d'un seul chercheur se retrouvent dans la version préliminaire.
- Le deuxième avantage permet de vérifier la justesse de la traduction par la comparaison de cette dernière avec la version originale de l'instrument. Le devient alors relativement facile de déceler les lacunes découlant de la traduction de l'instrument.

Cette méthode consiste à traduire la version originale du questionnaire qui est en anglais par un traducteur bilingue vers la langue cible qui est le français. Par la suite, la version française sera remise à un second traducteur bilingue qui la traduit en anglais. Cette dernière traduction doit se faire sans l'aide de la version originale du questionnaire, il s'agit d'essayer de reproduire la version originale du questionnaire en se servant uniquement de la version française. La justesse de cette dernière version s'évalue par le degré avec lequel elle aura permis de reproduire fidèlement la version originale. Il est proposé que ce jugement soit effectué lors de la phase d'évaluation de la version préliminaire réalisée par le comité.

Afin d'éviter certains partis pris, idées préconçues, interprétations et préjugés



qui seraient causés par des caractéristiques personnelles des deux premiers individus choisis pour réaliser la traduction inversée, il est parfois recommandé d'utiliser deux traductions inversées en parallèle. Deux traductions de l'anglais vers le français sont obtenues ainsi que deux traductions du français vers l'anglais. Les traductions de l'anglais vers le français représentent les versions préliminaires qui seront évaluées lors de la deuxième étape de la validation et les traductions du français vers l'anglais servent de moyens de comparaison avec l'instrument original.

Nous avons suivi la méthode de traduction inversée pour traduire la version anglaise de l'instrument D-RECT, en soumettant le questionnaire à un médecin bilingue pour le traduire en français. La traduction obtenue a été soumise après à un traducteur professionnel bilingue qui l'a traduite du français en anglais.

## **2. L'évaluation et la modification de la version préliminaire**

La version préliminaire obtenue lors de la phase de traduction devrait subir une évaluation minutieuse et cette phase constitue une étape importante lors du processus de validation du questionnaire. La méthode utilisée pour permettre cette évaluation et de composer un comité dont les membres seront habilités à effectuer une évaluation précise et objective de la version préliminaire(43,56). Il est préférable que le comité se compose de l'auteur de l'instrument original, les personnes qui ont développé la traduction, et les chercheurs pour obtenir différentes opinions et pour comprendre pourquoi certains termes anglais ont été traduits ou traduits de telle ou telle façon. L'évaluation des versions préliminaires se fait en deux étapes :

- La première étape consiste à comparer les différents items issus de la traduction inversée en anglais avec les items de la version originale anglaise. Lorsque les éléments provenant des traductions inversées et de la version anglaise originale sont identiques, les items en français ont alors satisfait un premier critère. En effet, ils ont permis un retour fidèle en langue originale et ceci représente un premier prérequis.
- La deuxième étape consiste à l'étude des termes utilisés dans la traduction française reliée au contenu psychologique. Ceci est très important. En effet, vu le vocabulaire nuancé de la langue française, il est possible que deux termes, intimement reliés, mais légèrement différents, permettent des traductions inversées identiques. Ceci explique pourquoi les termes français, même s'ils ont permis une traduction inversée fidèle, doivent faire l'objet d'une étude minutieuse par des experts-linguistes et par les chercheurs. De cette façon, les différences d'intensité ou d'expression des énoncés pourront être respectées.

Il'est à noter que si la forme française de certains éléments n'est pas satisfaisante, le comité se réserve alors le droit d'effectuer les changements jugés nécessaires. Cependant, ces changements, afin d'être incorporés dans le questionnaire français, devront respecter quatre règles bien précises(56)(43). La première règle est la suivante : (1) lorsque les traductions littérales ne sont pas possibles, la version expérimentale doit tenter de dégager le sens de l'élément original et non le « mot pour mot ». La deuxième règle (2) spécifie que les propriétés psycholinguistiques uniques de la langue cible doivent être utilisées autant que possible dans la formulation de la version expérimentale. En effet, ne pas tenir compte de ce point pourrait résulter en l'élaboration d'un instrument présenté dans la langue cible, mais n'ayant que très peu de sens pour les membres de la population qui sont ciblés. La troisième règle(3)

précise que si une mésentente survient entre les membres du comité en ce qui a trait à l'interprétation de certains éléments, des formes alternatives de ces éléments devront être incluses dans la traduction. La quatrième règle (4) précise que la version traduite française doit offrir le même format de présentation et les mêmes modalités de passation que ceux utilisés dans la version originale, car des présentations matérielles différentes du même instrument peuvent mener à des résultats différents(43,57).

Après cette deuxième étape, une version dite expérimentale de l'instrument choisi devrait être prête. Bien que certains chercheurs se limitent à cette seconde étape de validation, il est important de souligner que, d'après Vallerand(43), s'arrêter ici ne garantit pas la validité et la fidélité de l'instrument. Plus encore, s'arrêter après la seconde étape ne permet pas de juger du contenu du questionnaire tel qu'il peut apparaître aux yeux des sujets. Plusieurs recherches menées ont démontré que la perception de l'expérimentateur et celle du sujet ne concordent pas toujours(58,59)(43).

Dans notre travail, le comité était composé des personnes qui ont participé à la phase de traduction ainsi que les responsables du travail de la validation. Nous avons veillé à l'application des deux règles de l'évaluation pour aboutir à une version expérimentale pour la prochaine étape. Des arbitrages ont été faits pour adapter la traduction de certains termes au contexte du résidanat au Maroc. Deux exemples pour l'illustrer :

- Nous avons traduit le terme «Attending» en «Encadrant» en mettant entre parenthèses le mot «Enseignant», car la prise en charge des résidents est assurée essentiellement par les enseignants dans notre contexte.
- Un autre exemple. Le terme «Speciality tutor» a été traduit par «Chef de service», car dans notre contexte, la formation des résidents est sous la responsabilité du chef de service.

### **3. Le prétest(43)**

Le but du prétest, effectué dans un échantillon de la population cible, est de déterminer si les éléments qui composent la version expérimentale du questionnaire sont clairs, rédigés sans ambiguïté et dans un langage qui s'apparente bien à celui de la population cible. Les éléments qui sont déterminés ambigus doivent être examinés par les membres du comité et ceux-ci doivent les remplacer par des éléments qui sont porteurs du sens désiré.

Notre version expérimentale du questionnaire a été distribuée à 10 résidents des Centres hospitaliers universitaires de la ville Fès et Rabat. Les divers informations et retours obtenus ont facilité notre tâche dans la reformulation linguistique des items problématiques. Nous avons continué la phase de prétest jusqu'au moment où tous les items ont été jugés clairs.

### **4. Évaluation de la validité concomitante et de contenu**

La validité de contenu (content validity) réfère au jugement subjectif, et non quantitatif. Elle sert à démontrer que l'instrument mesure effectivement ce qu'il est censé mesurer(60). La validité concomitante (concurrent validity) par contre est démontrée empiriquement lorsqu'un test est fortement corrélé avec un critère mesurant le ou les mêmes concepts(61). Dans tous les cas, la détermination de la validité d'un instrument ou d'une mesure est le résultat d'un jugement pragmatique.

En ce qui concerne la traduction d'instruments psychologiques(43,62), dans la mesure où le chercheur démontre empiriquement que la version en langue française est étroitement reliée à la version originale, il devient alors possible de conférer à la fois validité concomitante et validité de contenu à l'instrument. Un moyen pratique de vérifier la validité est le recours à des sujets provenant d'une population représentative bilingue(43). Il s'agit donc de soumettre la forme originale et la version expérimentale

à des sujets bilingues provenant d'une population représentative afin d'étudier les coefficients de corrélation entre les résultats obtenus lors de la passation des deux versions. Les coefficients de corrélation peuvent être effectués soit entre le pointage total de la version expérimentale et de la version originale, soit entre chacun des éléments composant les instruments. Des coefficients de corrélation positifs élevés (plus de 0,75) indiquent qu'il existe bien une congruence entre les deux versions(43). Ceci permet d'avancer que la version expérimentale de l'instrument mesure les mêmes dimensions que celles de l'instrument original. Très peu d'études ont utilisé cette méthode dans leur validation de questionnaire, et nous n'avons pas utilisé cette méthode dans notre travail, car il est difficile dans notre contexte (essentiellement francophone) d'obtenir une large population parfaitement bilingue en anglais et en français).

## **5. Administration du questionnaire**

Dans l'étude d'élaboration du D-RECT menée par Boor et al.(28), un e-mail a été envoyé à tous les résidents des Pays-Bas, leur demandant de remplir le formulaire D-RECT. Pour les motiver à remplir le formulaire, une compensation financière pouvait être gagnée par les résidents qui remplissent le formulaire. Aucun rappel d'e-mail n'a été envoyé après. Lors de l'élaboration de la version courte du D-RECT par Silkens et al.(8), une version électronique et sur papier du formulaire a été distribuée en fonction des services. Dans le cas de la version électronique, un rappel a été envoyé 3 fois aux résidents, alors que pour la version papier aucun rappel n'a été fait.

Dans notre travail, nous avons préparé une version électronique du formulaire sur Google Form. Nous n'avons pas réussi à trouver une base de données d'e-mails fiable des résidents dans les facultés de médecine. Nous avons remédié à cette situation en désignant des résidents-référents dans chaque CHU qui se sont chargés de diffuser le lien électronique de notre formulaire auprès des résidents tout en leur

précisant de nous indiquer le nombre total de personnes contactées.

Dans la première semaine, nous avons reçu des retours des résidents signalant que notre formulaire contenait des éléments qui permettraient de les identifier par leur chef de service. Après concertation, le comité a décidé d'enlever la question qui précisait la spécialité des résidents.

Le recueil des données a été fait de manière anonyme et s'est étendu du 1er juillet au 30 septembre 2018. Le nombre de personnes qui a répondu à notre formulaire était de 211 par rapport à 695 de résidents contactés.

## **6. Analyse de la fidélité**

La fidélité d'un test psychologique concerne la précision de l'instrument, peu importe ce qu'il mesure(63). Un test fidèle mesure toujours le construit psychologique de la même façon. Les recherches récentes ont souligné l'utilisation de deux types d'indices : les indices de consistance interne ou homogénéité, et les indices de stabilité temporelle.

Lorsque le test est mesuré sur une échelle continue (comme une échelle de style Likert qui est le cas de la plupart des échelles contemporaines), l'analyse de l'alpha Cronbach test recommandée(43). Cette formule permet d'évaluer la fidélité d'un test qui aurait été administré une seule fois à une population d'individus. L'analyse va permettre d'étudier la corrélation des items entre eux et à quel point ils sont homogènes, tout en tenant compte du nombre d'items dans le test en question. La valeur du coefficient alpha peut varier entre 0 et 1. Plus les pointages sont élevés, c'est-à-dire plus ils sont rapprochés du nombre 1, plus l'instrument est jugé comme possédant un niveau élevé de consistance interne. Par contre, un pointage trop élevé (plus de 0,95 par exemple) indiquerait la présence d'une certaine redondance dans les éléments et peut laisser supposer que certains de ceux-ci mesurent un aspect trop restreint de la dimension visée. Des valeurs variantes entre 0,70 et 0,85 sont donc

généralement préférables(43,64).

Nous avons comparé les résultats de la consistance interne dans notre travail avec les travaux de Pacifico et al.(37), Boor et al.(28,37) et Silkens et al.(8) (Tableau 11). Comme pour les autres études, les scores obtenus varient entre 0,7 et 0,91, indiquant une bonne consistance interne de la traduction française.

**Tableau 11 : Résultats de la consistance interne des études de validation du D-RECT**

	Score Cronbach			
	Ce Travail	Pacifico et al.	Silkens et al.	Boor et al.
<b>Atmosphère éducative</b>	0,82	0,85	0,83	0,77
<b>Travail d'équipe</b>	0,79	0,93	0,79	0,69
<b>Rôle du chef de service</b>	0,91	0,96	0,86	0,78
<b>Coaching</b>	0,91	0,92	0,82	0,80
<b>Enseignement</b>	0,89	0,88	0,79	0,75
<b>Collaboration entre résidents</b>	0,81	0,92	0,84	0,76
<b>Travail adapté aux résidents</b>	0,70	0,88	0,71	0,66
<b>Disponibilité des encadrants</b>	0,82	0,93	0,71	0,64
<b>Sortie</b>	0,88	0,91	0,78	0,75

Les indices de stabilité temporelle s'évaluent en demandant à des sujets représentatifs de la population cible de remplir l'instrument à deux reprises. De plus, un écart de temps relativement important entre les deux passations de l'instrument doit être respecté. Si l'instrument possède cet aspect de la fidélité, les résultats obtenus lors des deux passations devraient demeurer stables. Des coefficients de corrélation positifs test-retest de l'ordre de 0,60 et plus, pour une telle période, s'avèrent satisfaisants(43). Nous avons demandé à 15 résidents de remplir le questionnaire deux fois, mais seuls 13 ont répondu. Le coefficient de corrélation a été 0,89, témoignant une bonne stabilité temporelle. La stabilité temporelle n'a pas été évaluée dans les précédents travaux de recherche de création ou de validation du D-RECT.

## **7. Validité du construit(52)(43,51)**

L'analyse factorielle confirmatoire est la méthode recommandée pour valider la structure factorielle d'un questionnaire. Lors de l'élaboration d'un nouvel instrument en sciences sociales, la structure factorielle est déterminée par l'analyse factorielle exploratoire, puis la structure retrouvée est validée par l'analyse factorielle confirmatoire. Dans le cas d'une validation d'une traduction d'un instrument, on utilise directement l'analyse factorielle confirmatoire en prenant comme modèle la structure proposée de l'instrument original. Bien que l'AFC soit fortement recommandée, il est intéressant de signaler qu'elle n'est pas systématiquement utilisée pour la validation d'un instrument. Ainsi, Pinnock et al. se sont limités à l'analyse de la consistance interne pour valider l'adaptation du D-RECT dans le contexte australien(35). De même, Caron et al.(23) ont uniquement utilisé la consistance interne (Cronbach) pour valider la traduction française du PHEEM (Postgraduate Hospital Educational Environment Measure).



### ➤ **Prérequis pour l'AFC**

Pour pouvoir utiliser l'AFC pour la validation du construit, il y a deux prérequis importants à respecter : normalité de la distribution de réponses et un nombre de participants suffisant (52).

Comme nous l'avons vu lors du chapitre matériel et méthode, Kline considère des valeurs absolues d'asymétrie inférieure à 3 et de Kurtosis inférieure à 10 comme acceptables pour réaliser une analyse factorielle confirmatoire. Cette condition est respectée dans cette étude.

### ➤ **Taille de l'échantillon**

Lors de la conception d'une enquête AFC, le chercheur doit répondre à la question cruciale de savoir combien de cas (participants) doivent être collectés pour obtenir un niveau acceptable de précision et de puissance statistique des estimations de paramètres du modèle, ainsi que des indices fiables de l'ajustement global du modèle. La littérature existante fournit peu de conseils sur cette question. Bien que les chercheurs s'accordent pour dire que plus la taille de l'échantillon est grande, mieux c'est pour l'AFC, il n'y a pas d'accord universel sur la taille suffisante. Malheureusement, il n'existe pas de moyen facile de déterminer la taille de l'échantillon nécessaire pour l'AFC. Les principales approches existantes sont : règles de base, méthode de Satorra— Saris et études de Monte-Carlo.

Certains recueils fournissent des règles empiriques générales basées sur un petit ensemble d'études. Il existe des directives très approximatives pour la taille des échantillons : moins de 100 est considéré comme «petit» et peut ne convenir que pour des modèles très simples; 100 à 200 est «moyen» et peut constituer une taille d'échantillon minimale acceptable si le modèle n'est pas trop complexe; et plus de 200 est «grand», ce qui est probablement acceptable pour la plupart des modèles(47,52). Des échantillons de petite taille peuvent entraîner des problèmes

techniques lors de l'exécution de l'analyse (comme des solutions non convergentes ou inappropriées) et une faible puissance(47). D'autres règles comme un nombre minimal de cas pour chaque paramètre libéré (par exemple, au moins 5 à 10 cas par paramètre) et nombre minimal de cas par indicateur dans le modèle (65,66).

La méthode Satorra - Saris(67) est une approche quantitative fondée sur un modèle, meilleure que les règles empiriques, mais présentant plusieurs inconvénients, elle n'est donc pas recommandée(51).

Muthén et Muthén(68) ont démontré l'utilisation d'une étude de Monte-Carlo avec Mplus (Logiciel statistique) pour identifier la taille d'échantillon appropriée et déterminer la puissance de l'AFC. Les études de Monte-Carlo sont conduites en générant des données à partir d'une population avec des valeurs de paramètres hypothétiques. Plusieurs échantillons sont prélevés, le modèle est estimé pour chaque échantillon et «les valeurs des paramètres et les erreurs types sont moyennées sur les échantillons».

L'échantillon dans notre étude a été de 211, ce qui répond aux exigences des règles empiriques sus-citées. Nous n'avons pas utilisé les autres méthodes de Sattora et Saris ou de Monte-Carlo.

#### ➤ **Interprétation des indices d'ajustement du modèle**

Il existe plusieurs indices de qualité d'ajustement, et la plupart d'entre eux peuvent être interprétés comme décrivant le manque d'ajustement du modèle aux données(52). Chaque type d'indice d'ajustement fournit des informations différentes sur l'ajustement du modèle (ou le non-ajustement), de sorte que les chercheurs indiquent généralement plusieurs indices d'ajustement lors de l'évaluation de l'ajustement du modèle.

Il existe de nombreuses directives pour un ajustement de modèle « acceptable » (47,69) :

- Brown recommande une RMSEA proche de 0,06 ou moins ; un SRMR proche de 0,08, un CFI proche de 0,95 ou plus ; et un TLI proche de 0,95 ou plus. Il est important de noter que ce ne sont pas des directives rigides, et Brown commente que son utilisation de « **prochede** » est intentionnelle.
- Kline(47) recommande de rapporter le modèle Khi-deux, le RMSEA, un intervalle de confiance de 90 % pour RMSEA, le CFI et SRMR. Selon Kline, un  $RMSEA \leq 0,05$  indique un ajustement approximatif proche, les valeurs comprises entre 0,05 et 0,08 suggèrent une erreur d'approximation raisonnable, et  $RMSEA \geq 0,10$  indique un ajustement médiocre. Un CFI « supérieur à environ 0,90 peut indiquer un ajustement relativement correct du modèle du chercheur », et les valeurs de SRMR « moins de 0,10 sont généralement considérées comme favorables ».

Il est important de noter que bien que Brown(47,51) et Kline(47) recommandent de rapporter plusieurs des mêmes indices d'ajustement, leurs critères d'ajustement acceptable sont différents, Brown étant un peu plus conservateur. Nous discuterons nos résultats par rapport aux recommandations de ces deux auteurs et des travaux publiés dans la littérature.

Le premier modèle testé a été le modèle original du questionnaire D-RECT proposé par Silkens et al.(8,47). Nous avons trouvé que le SRMR (Indice absolu), et RMSEA (indice de parcimonie) étaient adéquats selon les critères de Brown et Kline. Le CFI et le TLI (indices de comparaison) étaient légèrement inférieurs du seuil 0.90 proposé par Kline. Silkens et al.(8) ont rapporté de meilleurs résultats d'ajustement avec un SRMR et RMSEA de 0,04 (résultat bon), un CFI et TLI à 0,92 et 0,91 respectivement, ce qui est considéré acceptable selon Kline et Brown. Boor et al.(28)

ont obtenu un CFI à 0,89 (en dessous du seuil acceptable) et un RMSEA de 0,04 (bon) et a considéré ce résultat comme indiquant un bon ajustement du modèle.

Nous avons trouvé une seule étude qui a tenté de valider l'instrument D-RECT dans un contexte asiatique(28,37). Les auteurs n'ont pas obtenu un bon ajustement du modèle proposé par Silkens et al.(8). Par contre, ils ont proposé un modèle alternatif à 28 questions qui a donné de meilleurs résultats d'ajustement de modèle. Pour faire ce changement, ils se sont basés sur les indices de modifications pour éliminer les 7 questions du questionnaire. La valeur de cet index représente la baisse attendue de la valeur globale du Khi-deux lorsque le paramètre doit être estimé librement. Ainsi, l'indice de modification indique le gain d'ajustement potentiel lors de la suppression d'un élément du modèle(37).

Vu nos résultats, et comparativement aux résultats obtenus par Boor et Silkens, nous pouvons considérer que nous avons obtenu un ajustement de modèle adéquat. Sur les trois types d'indices, deux types d'indices étaient satisfaisants (absolu et parcimonie), alors que les indices de comparaison étaient proches de la valeur seuil (0,89). Le même résultat a été trouvé par Boor lors de la création du questionnaire D-RECT original (CFI = 0,89) et il a considéré le modèle satisfaisant, et ses conclusions ont été acceptées par la communauté scientifique.

Nous avons décidé d'explorer la possibilité d'améliorer l'ajustement du modèle. Dans le chapitre suivant, nous discuterons cette procédure, en expliquant ce qui est recommandé dans la littérature.

#### ➤ Révision du modèle(37,51,52)

Bien que les résultats obtenus sont satisfaisants, nous avons décidé d'explorer la révision du modèle. Notre principale motivation est pédagogique. La validation psychométrique de questionnaires est une procédure complexe et il n'existe pas de références francophones récentes pour guider le chercheur. Nous avons voulu aider les

chercheurs en expliquant au mieux toutes les étapes de cette procédure complexe, en orientant vers les références anglophones les plus pertinentes.

Si un modèle est mal ajusté, le chercheur devra identifier les causes de mauvais ajustement; ensuite, en fonction des zones de mauvais ajustement et des révisions indiquées, le chercheur peut modifier le modèle. Il existe de nombreuses façons d'identifier les zones de mauvais ajustement, mais le modèle ne doit être modifié que s'il est cohérent avec la théorie ou les recherches antérieures et s'il est «logique». Les zones de mauvais ajustement peuvent être identifiées en examinant les indices de modification et les zones localisées de contrainte (c.-à-d. résidus).

Il est important de noter qu'une fois on a commencé à modifier un modèle basé sur les IM ou des résidus standardisés, même si les modifications sont justifiées, le travail sort du cadre de confirmation pour passer à un travail exploratoire. Par conséquent, les modèles respécifiés doivent être interprétés avec prudence et les modifications substantielles doivent être répliquées dans des échantillons indépendants ou, si l'évaluation initiale et la révision du modèle ont une taille d'échantillon suffisamment grande, elles peuvent être fractionnées de manière aléatoire en deux afin que l'AFC initiale par le modèle puisse être effectuée sur la moitié de l'échantillon, puis que l'ajustement du modèle final puisse être testé dans la seconde moitié de l'échantillon.

#### ➤ **Indices de modification**

Les indices de modification (IM) sont générés par les progiciels (Ex. SPSS Amos). Ce sont des indicateurs basés sur les données des changements apportés au modèle et susceptibles d'améliorer leur ajustement. Un IM supérieur à 3,84 (ou environ 4) indique un changement qui entraînera probablement une amélioration significative de l'ajustement du modèle(52). L'IM peut suggérer des modifications à n'importe quel aspect du modèle, notamment en ajoutant des chemins entre variables latentes, en

ajoutant des chemins de variables latentes à des variables observées non spécifiées à l'origine comme indicateurs de cette variable latente, en ajoutant des covariances d'erreur entre les variables observées, etc. Les IM pour les covariances suggère d'ajouter des covariances d'erreur, soit entre deux erreurs, soit entre une erreur et une variable latente. L'IM pour les variances suggère d'ajouter des variances entre les variables latentes. L'IM pour les poids de régression suggère d'ajouter des chemins de régression au modèle. Par exemple, les chemins suggérés peuvent aller d'une variable latente à une variable observée ou d'une variable observée à une autre. Bon nombre des modifications suggérées par les IM peuvent ne pas avoir de sens compte tenu de la théorie et des recherches antérieures ; de telles modifications non sensées ne doivent pas être faites, quelle que soit l'ampleur du changement de paramètre.

Dans notre cas, en observant les IM, nous avons noté des IM élevés entre les variables observées. Nous avons ajouté des covariances d'erreur entre des variables observées appartenant à la même variable latente (ou facteur) et qui avaient des IM supérieurs à 10. Cela a permis d'améliorer le modèle, en obtenant de meilleurs résultats sur tous les indices (voir tableau 10).

#### ➤ **Zones localisées de contrainte(52)**

Les résidus peuvent être examinés pour identifier les zones localisées de contrainte. Les résidus standardisés supérieurs à 1,96 (pour  $p < 0,05$ ) ou à 2,58 (pour  $p < 0,01$ ) peuvent indiquer des zones de contrainte. Les résidus standardisés positifs indiquent que les paramètres du modèle sous-estiment la relation, tandis que les résidus standardisés négatifs indiquent que les paramètres du modèle surestiment la relation. Nous n'avons pas exploré les zones localisées de contraintes dans notre étude.

### ➤ Modèles imbriqués

La révision du modèle peut aboutir à un modèle imbriqué «contenant un sous-ensemble des paramètres libres d'un autre modèle, souvent appelé modèle parent» (Brown, 2006, p. 48). Le test de différence  $\chi^2$  (Khi-deux) peut être utilisé pour tester la significativité d'une amélioration de modèle avec des modèles imbriqués. Le test de différence  $\chi^2$  est effectué en déterminant la différence entre  $\chi^2$  pour les modèles parent et imbriqué (c.-à-d. 2 pour le modèle parent moins  $\chi^2$  pour le modèle imbriqué), puis en déterminant la différence entre les degrés de liberté pour le modèle parent et modèles imbriqués (c'est-à-dire, degrés de liberté pour le modèle parent moins df pour le modèle imbriqué). La différence  $\chi^2$  est ensuite testée pour déterminer la signification de la différence du degré de liberté associée.

Dans notre exemple, le  $\chi^2$  du modèle parent est égal à 1126228 et le  $\chi^2$  du modèle imbriqué est égal à 1003197. La différence  $\chi^2$  est égale à 123031. Le degré de liberté du modèle parent est égal à 524 et celui du modèle imbriqué est égal à 518, la différence df est égale à 6. En utilisant un tableau de valeurs critiques pour  $\chi^2$ , nous constatons que pour un degré de liberté de 6; la valeur critique supérieure est égale à 12592, et la valeur critique inférieure est 1635 (voir <http://www.itl.nist.gov/div898/handbook/eda/section3/eda3674.htm>). Par conséquent, comme notre différence  $\chi^2$  obtenue est située entre ces intervalles, nous pouvons en conclure que le changement de modèle n'a pas entraîné une amélioration significative ( $p > 0,05$ ) de l'ajustement du modèle.

## **B. Points forts**

Le principal point fort du travail est qu'il a permis de valider la version française de l'instrument D-RECT, permettant ainsi son utilisation dans la recherche scientifique, non seulement dans notre pays, mais aussi dans les autres pays francophones.

L'un des enjeux de la recherche médicale est la reproductibilité des résultats scientifiques obtenus. Dans ce cas, la reproductibilité des résultats obtenus par les auteurs originaux suggère la solidité de l'instrument D-RECT et son adaptabilité à d'autres systèmes de résidanat internationaux.

L'autre point fort de ce travail a été la participation de résidents représentant les cinq centres hospitaliers universitaires nationaux à cette étude. Une autre force de ce travail a été que l'échantillon qui a participé était représentatif de toutes les catégories de résidents dans le système national : résidents sur concours, résidents sur titre, bénévoles, contractuels et même des médecins étrangers.



## C. Limites du travail

Une des limites de ce travail a relativement été le faible effectif de notre échantillon comparativement aux études précédentes. Lors de l'élaboration de la première version de l'instrument D-RECT par Boor et al., 1276 résidents ont participé à l'étude de validation(28). Lors de l'élaboration de la version modifiée par Silkens et al., 2306 résidents ont participé au travail(8,28). Finalement, 843 résidents ont participé au travail de validation du D-RECT aux Philippines(37). Ces auteurs ont pu atteindre ce nombre important, car les investigateurs avaient accès aux contacts de tous les résidents de leurs pays. Aux Pays-Bas, le D-RECT est utilisé régulièrement à échelle institutionnelle pour évaluer l'environnement de formation des résidents dans le pays, et donc, Silkens et al. avaient accès aux réponses de tous les résidents nationaux. Finalement, aux Philippines, Pacifico et al. ont administré le questionnaire lors du passage d'un examen national pour les résidents de deuxième et troisième année en médecine interne.

Au Maroc, les facultés de médecine n'ont pas une base de données des mails des résidents en formation et il était nécessaire de faire du porte-à-porte pour distribuer le lien du formulaire aux résidents. Ceci pourrait expliquer le faible nombre de participations, comparativement aux autres études. Néanmoins, il faut noter que c'est relatif vu que notre effectif est suffisant pour réaliser une analyse factorielle confirmatoire, comme nous l'avons expliqué précédemment.

L'autre limite de ce travail a été l'absence d'identification des services cliniques où se forment les résidents. Nous avons enlevé la question du formulaire, sur demande des résidents, qui avaient exprimé la peur d'être identifiés par leurs encadrants. De ce fait, il ne nous a pas été possible de réaliser une étude de généralisabilité pour estimer le nombre minimum de résidents par service, nécessaires pour une évaluation fiable de l'environnement de formation.

Finalement, nous n'avons pas discuté ou analysé les résultats du D-RECT obtenus dans notre échantillon, car l'objectif principal de notre travail a été la validation de l'instrument et non l'évaluation de l'environnement de formation.

## **D. Implications et futures pistes de recherche**

La validation de la version française de D-RECT dans le contexte national ouvre des perspectives intéressantes pour la recherche et l'évaluation de la perception des environnements de formation par les résidents.

La perspective la plus intéressante est certainement l'utilisation de l'instrument dans le cadre d'initiatives d'amélioration de la qualité. Par exemple, Le D-RECT est largement utilisé aux Pays-Bas pour surveiller en permanence l'environnement de formation perçu par les résidents. Le D-RECT est ici utilisé avec l'objectif d'identifier les domaines à améliorer et, à ce titre, de servir de base aux initiatives d'amélioration de la qualité(10,37). Ainsi, il a été montré une amélioration significative de la perception de l'environnement de formation, monitorée par le D-RECT. On pourrait donc proposer aux facultés de médecine nationales d'incorporer l'évaluation de l'environnement de formation aux démarches d'amélioration de la qualité locales. À petite échelle, les chefs de service pourraient utiliser cet instrument dans leurs unités pour les mêmes objectifs. Ceci permettrait d'identifier les domaines de l'environnement de formation à améliorer et déployer des stratégies spécifiques de correction.

Cette validation ouvre également des perspectives de recherche en pédagogie médicale. L'importance de l'environnement de formation a été clairement démontrée dans plusieurs travaux scientifiques internationaux(10,40,70,71) et il serait intéressant d'analyser la corrélation entre l'environnement et les autres aspects de la formation des résidents au Maroc.

# CONCLUSION

L'amélioration de la qualité de formation des résidents est un enjeu majeur pour les facultés de médecine. L'environnement de formation constitue un élément important à prendre en considération par les institutions académiques, vu l'importance de son impact sur la qualité de formation et le bien-être des résidents.

Ce travail a permis de validation psychométrique de la traduction française du Dutch Residency Educational Climate Test. Nous avons montré que cette version française avait une bonne consistance interne, une bonne fiabilité temporelle et une validité du construit acceptable.

Cette validation permettra d'utiliser le D-RECT pour l'évaluation de la perception de l'environnement de formation par les résidents au Maroc et aussi dans les autres pays francophones.

L'utilisation du D-RECT ouvre également des perspectives intéressantes dans le cadre de démarche d'amélioration de la qualité des systèmes de formation spécialisée en médecine, ou dans le cadre de la recherche scientifique en pédagogie médicale au Maroc.

Finalement, nous avons décrit avec détail la procédure de validation psychométrique. Nous espérons que ça sera un guide utile pour les futurs chercheurs marocains pour mener des travaux de recherche similaires.

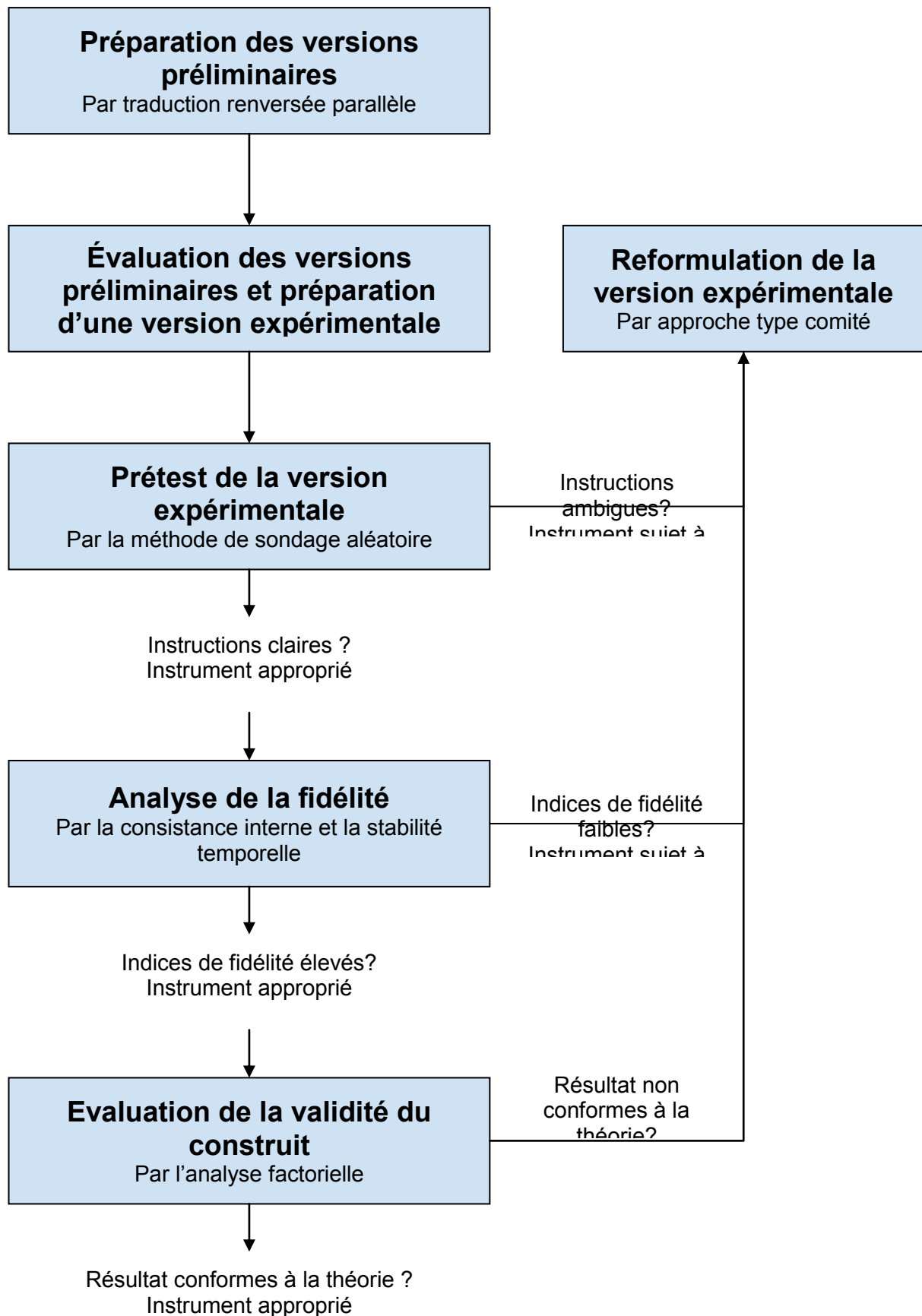


Figure 6 : Présentation séquentielle des différentes phases de la méthodologie de validation transculturelle de tests psychologiques. Selon Vallerand(43).

# RÉSUMÉS

# RÉSUMÉ

## Validation de la traduction française du « Dutch Residency Educational Climate Test » D-RECT

### Introduction

L'environnement de formation est l'un des aspects les plus influents dans la formation des résidents en médecine. Il existe plusieurs instruments pour évaluer l'environnement de formation dans la littérature, et le plus solide d'entre eux est le Dutch Residency Educational Climate Test (D-RECT). Cependant, aucun instrument n'a été validé dans le contexte marocain. L'objectif de ce travail est la validation psychométrique de la traduction française du D-RECT dans le contexte du résidanat de médecine au Maroc.

### Matériel et méthodes

Après traduction du questionnaire D-RECT en français, en respectant les recommandations internationales, les résidents des cinq centres hospitaliers marocains ont été invités à remplir le formulaire entre juillet et septembre 2018. La consistance interne, la stabilité temporelle et l'analyse factorielle confirmatoire ont été utilisées pour analyser les propriétés psychométriques de la version traduite.

### Résultats

Durant la période du travail, 211 résidents ont rempli le questionnaire. L'analyse de la fidélité de l'instrument a montré que cette version française avait une bonne consistance interne (Score de Cronbach alpha = 0,95) et une bonne stabilité temporelle (score de corrélation entre deux mesures = 0,89). L'analyse factorielle confirmatoire a montré que la traduction française a un ajustement de modèle adéquat avec les scores suivants sur les indicateurs d'ajustement : SRMS = 0,058/RMSEA = 0,07/CFI = 0,88/TLI = 0,87/Khi-deux normé = 2,15.

## **Conclusion**

Ce travail a permis la validation psychométrique de la traduction en français du D-RECT. Les chercheurs marocains disposent donc d'un outil validé pour étudier les environnements de formation au Maroc, et les décideurs peuvent l'utiliser dans le cadre d'une démarche d'amélioration des systèmes de résidanat.



# **ABSTRACT**

## **Validation of the French translation of the « Dutch Residency Educational Climat Test » D-RECT**

### **Introduction**

The educational environment is one of the most influential aspects in the training of medical residents. There are several instruments to evaluate the educational environment in the literature, and the strongest of these is the Dutch Residency Educational Climate Test (D-RECT). However, no instrument has been validated in the Moroccan context. The objective of this work is the psychometric validation of the French translation of D-RECT in the context of the medical residency in Morocco.

### **Material and Methods**

After translation of the D-RECT questionnaire into French, respecting the international recommendations, the residents of five Moroccan university hospitals were invited to complete the form between July and September 2018. Internal consistency, temporal stability and confirmatory factor analysis been used to analyze the psychometric properties of the translated version.

### **Results**

During the study period, 211 residents completed the questionnaire. The analysis of the fidelity of the instrument showed that this French version had a good internal consistency (Cronbach alpha score = 0,95) and a good temporal stability (correlation score between two measurements = 0.89). Confirmatory factor analysis showed that the French translation has an adequate model fit with the following scores on the adjustment indicators: SRMS = 0.058 / RMSEA = 0,07/CFI = 0,88/TLI = 0,87/ Normalized Chi-2 = 2.15.

## **Conclusion**

This work enabled the psychometric validation of the French translation of D-RECT. Moroccan researchers have a validated tool for studying training environments in Morocco, and decision-makers can use it as part of a quality improvement process of residency systems.

## ملخص

### التحقق من صحة الترجمة الفرنسية لـ

#### « Dutch Residency Educational Climat Test » D-RECT

##### مقدمة

بيئة التدريب هي واحدة من أكثر الجوانب المؤثرة في تدريب المقيمين الطبيين. هناك العديد من الأدوات لتقييم بيئة التدريب في الأدب العلمي ، وأقواها هو اختبار المناخ التربوي التعليمي في هولندا (D-RECT) ومع ذلك ، لم يتم التحقق من صحة أي أداة في السياق المغربي. الهدف من هذا العمل هو التحقق من الخصائص السيكومترية للترجمة الفرنسية لـ D-RECT في سياق الإقامة الطبية في المغرب.

##### المواد والطرق

بعد ترجمة D-RECT باللغة الفرنسية، وفقا للتوصيات الدولية، طلب من المقيمين الطبيين في خمسة مستشفيات جامعية مغربية ملء الاستمارة بين يوليو وسبتمبر 2018. والاتساق الداخلي والاستقرار الزمني والتحليل العاملي التوكيدي كانت تم استخدامها لتحليل الخصائص السيكومترية للنسخة المترجمة.

##### النتائج

خلال فترة العمل ، أكمل 211 من المقيمين الاستبيان. النتائج أظهرت أن هذه النسخة الفرنسية لها اتساق داخلي جيد (ألفا = 0.95 درجة كرونباخ) واستقرار زمني جيد (درجة الارتباط بين القياسات = 0.89). وأظهر التحليل العاملي التوكيدي أن الترجمة الفرنسية هو نموذج مناسب لنتناسب مع مؤشرات التعديل التالية  $SRMS = 0.058$  /  $RMSEA = 0,07$  /  $CFI = 0,88$  /  $0,87 = TLI$  /  $normed.$

##### استنتاج

ممكن هذا العمل من التحقق من الخصائص السيكومترية للترجمة الفرنسية لـ D-RECT. وبالتالي ، لدى الباحثين المغاربة أداة معتمدة لدراسة بيئات التدريب في المغرب ، ويمكن لصناع القرار استخدامها كجزء من عملية لتحسين أنظمة الإقامة الطبية في المغرب.

# **BIBLIOGRAPHIE**

- [1]. Boor K-B. The Clinical Learning Climate. 2009. 131 p.
- [2]. Genn JM. AMEE Medical Education Guide No. 23 (Part 2): Curriculum, environment, climate, quality and change in medical education – a unifying perspective. *Med Teach*. 2001;23(5):445-54.
- [3]. Rothman AI, Ayoade F. The development of a learning environment: a questionnaire for use in curriculum evaluation. *J Med Educ*. 1970 Oct; 45(10):754-9.
- [4]. Genn JM. AMEE Medical Education Guide No. 23 (Part 1): Curriculum, environment, climate, quality and change in medical education—a unifying perspective. *Med Teach*. 2001 Jul;23(4):337-44.
- [5]. Rothman AI, Ayoade F. The development of a learning environment. *Acad Med*. 1970;45(10):754-9.
- [6]. Pace CR, Robert Pace C, Stern GG. An approach to the measurement of psychological characteristics of college environments. *J Educ Psychol*. 1958;49(5):269-77.
- [7]. Genn JM, Harden RM. What is Medical Education Here Really Like? Suggestions for action research studies of climates of medical education environments. *Med Teach*. 1986;8(2):111-24.
- [8]. Silkens MEWM, Smirnova A, Stalmeijer RE, Arah OA, Scherpbier AJJA, Van Der Vleuten CPM, et al. Revisiting the D-RECT tool: Validation of an instrument measuring residents' learning climate perceptions. *Med Teach*. 2016 May;38(5):476-81.
- [9]. McAleer SRS, Roff S, McAleer S. What is educational climate? *Med Teach*. 2001;23(4):333-4.

- [10]. Silkens MEWM, Arah OA, Scherpbier AJJA, Heineman MJ, Lombarts KMJMH. Focus on Quality: Investigating Residents' Learning Climate Perceptions. *PLoS One*. 2016 Jan 14;11(1):e0147108.
- [11]. Wall D, Clapham M, Riquelme A, Vieira J, Cartmill R, Aspegren K, et al. Is PHEEM a multi-dimensional instrument? An international perspective. *Med Teach*. 2009 Nov;31(11):e521-7.
- [12]. Llera J, Durante E. Correlation between the educational environment and burn-out syndrome in residency programs at a university hospital . *Archivosargentinos de pediatría*. 2014 Feb 1;112(1):6-11.
- [13]. Gronlund NE, Linn RL. *Measurement and Evaluation in Teaching*. MacMillan Publishing Company; 1990. 530 p.
- [14]. Nunnally JC. *Psychometric theory*. McGraw-Hill Companies; 1978. 701 p.
- [15]. Soemantri D, Herrera C, Riquelme A. Measuring the educational environment in health professions studies: a systematic review. *Med Teach*. 2010;32(12):947-52.
- [16]. Nagraj S, Wall D, Jones E. The development and validation of the mini-surgical theater educational environment measure. *Med Teach*. 2007 Sep;29(6):e192-7.
- [17]. Kanashiro J, McAleer S, Roff S. Assessing the educational environment in the operating room—a measure of resident perception at one Canadian institution. *Surgery*. 2006 Feb;139(2):150-8.
- [18]. Holt MC, Roff S. Development and validation of the Anaesthetic Theatre Educational Environment Measure (ATEEM). *Med Teach*. 2004 Sep;26(6):553-8.
- [19]. Mulrooney A. Development of an instrument to measure the Practice Vocational Training Environment in Ireland. *Med Teach*. 2005 Jun 3;27(4):338-42.

- [20]. Roth LM, Severson RK, Probst JC, Monsur JC, Markova T, Kushner SA, et al. Exploring physician and staff perceptions of the learning environment in ambulatory residency clinics. *Fam Med*. 2006 Mar;38(3):177-84.
- [21]. Keitz SA, Holland GJ, Melander EH, Bosworth HB, Pincus SH, VA Learners' Perceptions Working Group. The Veterans Affairs Learners' Perceptions Survey: the foundation for educational quality improvement. *Acad Med*. 2003 Sep;78(9):910-7.
- [22]. Bassaw B, Roff S, McAleer S, Roopnarinesingh S, De Lisle J, Teelucksingh S, et al. Students' perspectives on the educational environment, Faculty of Medical Sciences, Trinidad. *Med Teach*. 2003 Sep; 25(5) : 522-6.
- [23]. Caron F, Pina A, Mahone M, Costa J-P, Sansregret A, Durand M. Évaluer l'environnement éducatif post-gradué : traduction et validation d'un questionnaire. *Pédagogiemédicale*. 2014 May 18;15(2):91-8.
- [24]. Aspegren K, Bastholt L, Basted KM, Bonnesen T, Ejlersen E, Fog I, et al. Validation of the PHEEM instrument in a Danish hospital setting. *Med Teach*. 2007 Jun;29(5):498-500.
- [25]. Riquelme A, Herrera C, Aranis C, Oporto J, Padilla O. Psychometric analyses and internal consistency of the PHEEM questionnaire to measure the clinical learning environment in the clerkship of a Medical School in Chile. *Med Teach*. 2009 Jun;31(6):e221-5.
- [26]. Boor K, Scheele F, van der Vleuten CPM, Scherpbier AJJA, Teunissen PW, Sijtsma K. Psychometric properties of an instrument to measure the clinical learning environment. *Med Educ*. 2007 Jan;41(1):92-9.
- [27]. Vieira JE. The postgraduate hospital educational environment measure (PHEEM) questionnaire identifies quality of instruction as a key factor predicting academic achievement. *Clinics*. 2008 Dec;63(6):741-6.

- [28]. Boor K, Van Der Vleuten C, Teunissen P, Scherpbier A, Scheele F. Development and analysis of D-RECT, an instrument measuring residents' learning climate. *Med Teach*. 2011 Feb 28;33(10):820-7.
- [29]. Roff S, McAleer S, Harden RM, Al-Qahtani M, Ahmed AU, Deza H, et al. Development and validation of the Dundee Ready Education Environment Measure (DREEM). *Med Teach*. 1997 Jan 3;19(4):295-9.
- [30]. Schönrock-Adema J, Heijne-Penninga M, Van Hell EA, Cohen-Schotanus J. Necessary steps in factor analysis: enhancing validation studies of educational instruments. The PHEEM applied to clerks as an example. *Med Teach*. 2009 Jun;31(6):e226-32.
- [31]. Ramani S, Leinster S. AMEE Guide no. 34: Teaching in the clinical environment. *Med Teach*. 2008;30(4):347-64.
- [32]. Farnan JM, Petty LA, Georgitis E, Martin S, Chiu E, Prochaska M, et al. A systematic review: the effect of clinical supervision on patient and residency education outcomes. *Acad Med*. 2012 Apr;87(4):428-42.
- [33]. Schönrock-Adema J, Bouwkamp-Timmer T, van Hell EA, Cohen-Schotanus J. Key elements in assessing the educational environment: where is the theory? *Adv Health Sci Educ Theory Pract*. 2012 Dec;17(5):727-42.
- [34]. Ostroff C. The Effects of Climate and Personal Influences on Individual Behavior and Attitudes in Organizations. *Organ Behav Hum Decis Process*. 1993 Oct;56(1):56-90.
- [35]. Pinnock R, Welch P, Taylor-Evans H, Quirk F. Using the DRECT to assess the intern learning environment in Australia. *Med Teach*. 2013 Aug;35(8):699.
- [36]. Bennett D, Dornan T, Bergin C, Horgan M. Postgraduate training in Ireland: expectations and experience. *Ir J Med Sci*. 2014 Dec;183(4):611-20.



- [37]. Pacifico JL, van der Vleuten CPM, Muijtjens AMM, Sana EA, Heeneman S. Cross-validation of a learning climate instrument in a non-western postgraduate clinical environment. *BMC Med Educ.* 2018 Jan 25;18(1):22.
- [38]. Iblher P, Zupanic M, Ostermann T. The Questionnaire D-RECT German : Adaptation and testtheoretical properties of an instrument for evaluation of the learning climate in medical specialist training. *GMS Z Med Ausbild.* 2015 Nov 16;32(5):Doc55.
- [39]. Alshomrani AT, AlHadi AN. Learning environment of the Saudi psychiatry board training program. *Saudi Med J.* 2017 Jun;38(6):629–35.
- [40]. van Vendeloo SN, Godderis L, Brand PLP, Verheyen KCPM, Rowell SA, Hoekstra H. Resident burnout: evaluating the role of the learning environment. *BMC Med Educ.* 2018 Mar 27;18(1):54.
- [41]. Lombarts KMJMH, Heineman MJ, Scherpbier AJJA, Arah OA. Effect of the learning climate of residency programs on faculty's teaching performance as evaluated by residents. *PLoS One.* 2014 Jan 28;9(1):e86512.
- [42]. Silkens MEWM, Chahine S, Lombarts KMJMH, Arah OA. From good to excellent: Improving clinical departments' learning climate in residency training. *Med Teach.* 2018 Mar;40(3):237–43.
- [43]. Vallerand RJ. Vers une méthodologie de validation transculturelle de questionnaires psychologiques : Implications pour la recherche en langue française. *Canadian Psychology/Psychologie canadienne.* 1989; 30(4) : 662–80.
- [44]. Malhotra NK, Kim SS, Agarwal J. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research.* 2004 Dec;15(4):336–55.

- [45]. Pinsonneault A, Kraemer K. Survey Research Methodology in Management Information Systems: An Assessment. *Journal of Management Information Systems*. 1993 Sep 15;10(2):75-105.
- [46]. Fowler F. *Survey Research Methods* (4th ed.). 2455 Teller Road, Thousand Oaks California 91320 United States : SAGE Publications, Inc.; 2009.
- [47]. Kline T. *Psychological Testing: A Practical Approach to Design and Evaluation*. 2455 Teller Road, Thousand Oaks California 91320 United States : SAGE Publications, Inc.; 2005.
- [48]. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *J Int Assoc Med Sci Educ*. 2011 Jun 27;2:53-5.
- [49]. Tavakol M, Mohagheghi MA, Dennick R. Assessing the skills of surgical residents using simulation. *J Surg Educ*. 2008 Mar;65(2):77-83.
- [50]. Bland JM, Altman DG. Cronbach's alpha. *BMJ*. 1997 Feb 22;314(7080):572.
- [51]. Brown TA. *Confirmatory Factor Analysis for Applied Research*, Second Edition. Guilford Publications; 2015. 462 p.
- [52]. Harrington D. *Confirmatory Factor Analysis*. 2008.
- [53]. Pillay H, Irving K, McCrindle A. Developing a diagnostic tool for assessing tertiary students' readiness for online learning. *IJLT*. 2006;2(1):92.
- [54]. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 2007. 980 p.
- [55]. Brislin RW, Thorndike RM, Lonner WJ. *Cross-cultural Research Methods*: (by) Richard W. Brislin, Walter J. Lonner (and) Robert M. Thorndike. 1973. 352 p.
- [56]. Hambleton RK, Merenda PF, Spielberger CD. *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Psychology Press; 2004. 392 p.

- [57]. Converse JM, Presser S. Survey Questions: Handcrafting the Standardized Questionnaire. SAGE; 1986. 80 p.
- [58]. Russell D. The Causal Dimension Scale: A measure of how individuals perceive causes. *J Pers Soc Psychol.* 1982 ;42(6) : 1137-45.
- [59]. McAuley E, Gross JB. Perceptions of Causality In Sport: An Application of the Causal Dimension Scale. *Journal of Sport Psychology.* 1983;5(1):72-6.
- [60]. Anastasi A. Psychological testing. 1976. 750 p.
- [61]. Allen MJ, Yen WM. Introduction to Measurement Theory. Waveland Press; 2001. 310 p.
- [62]. Spielberger CD, Díaz-Guerrero R. Cross-cultural Anxiety. Halsted Press; 1976. 195 p.
- [63]. Nunnally (Jr. J. Introduction to Psychological Measurement. 1970. 572 p.
- [64]. Mclver J, Carmines E. Unidimensional Scaling. 1981.
- [65]. Bentler PM, Chou C-P. Practical Issues in Structural Modeling. *Sociol Methods Res.* 1987 Aug 30;16(1):78-117.
- [66]. Ding L, Velicer WF, Harlow LL. Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Struct Equ Modeling.* 1995 Jan;2(2):119-43.
- [67]. Satorra A, Saris WE. Power of the likelihood ratio test in covariance structure analysis. *Psychometrika.* 1985 Mar;50(1):83-90.
- [68]. Muthén LK, Muthén BO. How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power. *Struct Equ Modeling.* 2002 Oct;9(4):599-620.
- [69]. Raykov T, Tomer A, Nesselroade JR. Reporting structural equation modeling results in Psychology and Aging: Some proposed guidelines. *Psychol Aging.* 1991;6(4):499-503.

- [70]. Silkens MEWM, Lombarts KMJMH, Scherpbier AJJA, Heineman MJ, Arah OA. Towards healthy learning climates in postgraduate medical education: exploring the role of hospital-wide education committees. *BMC Med Educ.* 2017 Dec 6;17(1):241.
- [71]. Silkens MEWM, Arah OA, Wagner C, Scherpbier AJJA, Heineman MJ, Lombarts KMJMH. The Relationship Between the Learning and Patient Safety Climates of Clinical Departments and Residents' Patient Safety Behaviors. *Acad Med.* 2018 Sep;93(9):1374–80.