ciep

AU SERVICE
DU FRANÇAIS
ET DE L'ÉDUCATION
DANS LE MONDE

Impact de la méthode de calcul du score sur les résultats des candidats

Marianne Laurent & Sébastien GEORGES
Centre international d'études pédagogiques
Département évaluation et certifications
Cellule qualité et expertises - psychométrie

ciep
AU SERVICE DU FRANÇAIS ET DE L'ÉDUCATION DANS LE MONDE

Département
évaluation et certifications

ciep
AU SERVICE
DU FRANÇAIS
ET DE L'ÉDUCATION
DANS LE MONDE

Évaluation et certifications

Gestion pédagogique et administrative des certifications nationales pour publics non francophones

- DILF
- DELF et ses déclinaisons
  - DELF Prim
  - DELF scolaire et junior
  - DELF Pro
- DALF
- TCF et ses déclinaisons
  - TCF tout public
  - TCF DAP
  - TCF Québec
  - TCF ANF

Expertise et psychométrie : habilitations, appui à la création et au développement de certifications étrangères, audits, etc.

Test de connaissance du français

- Le TCF est le test de français du ministère de l'Éducation nationale, de l’enseignement supérieur et de la recherche.

- Le TCF est un test de niveau linguistique en français langue générale destiné aux personnes qui souhaitent, pour des raisons professionnelles, personnelles ou d’études, faire évaluer et valider leurs connaissances en français de façon simple, fiable et rapide.

- Le TCF est un test standardisé et calibré conçu par le CIEP, organisme certificateur membre de ALTE, selon une méthodologie rigoureuse qui en fait un véritable instrument de mesure.

Objectif

- Déterminer l’impact de différentes méthodes de calcul du score sur les résultats des candidats aux épreuves à questions à choix multiple du TCF.
  - Comparer l’actuelle méthode de calcul des scores du TCF à d’autres méthodes en :
    - mesurant l’impact de l’introduction de points négatifs dans le calcul du score brut
    - mesurant l’impact de l’application d’un critère proportionnel de réussite d’items

Méthodologie

ciep
AU SERVICE
DU FRANÇAIS
ET DE L'ÉDUCATION
DANS LE MONDE

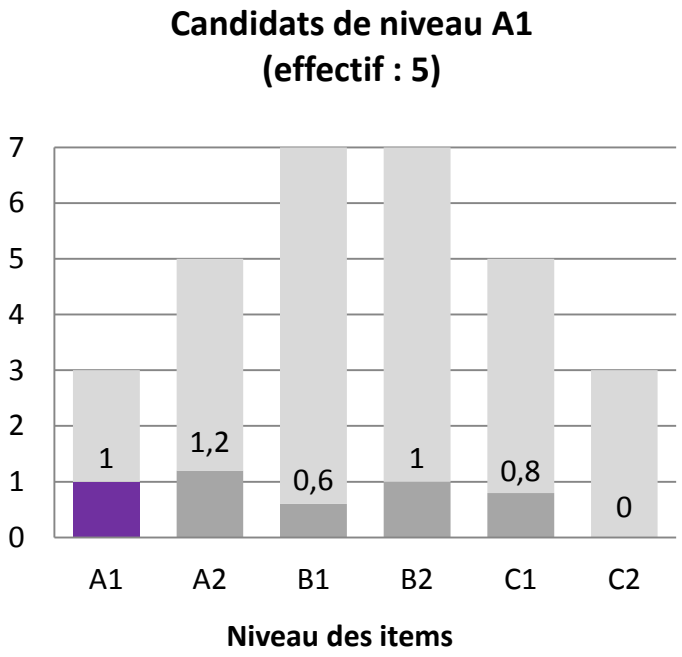
Echantillon

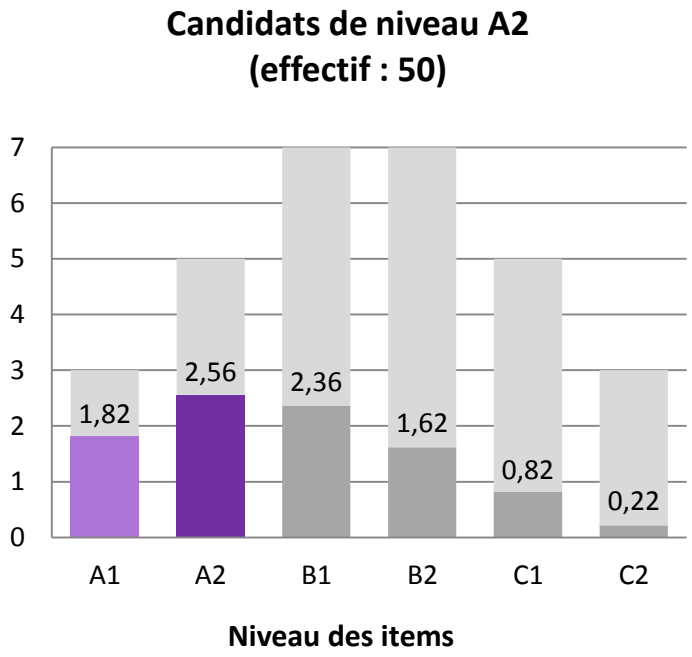
- Extraction des résultats :
  - n = 1 039 candidats (TCF tout public, premier trimestre 2012)
  - 83 120 données = matrice de 80 réponses x 1 039 candidats

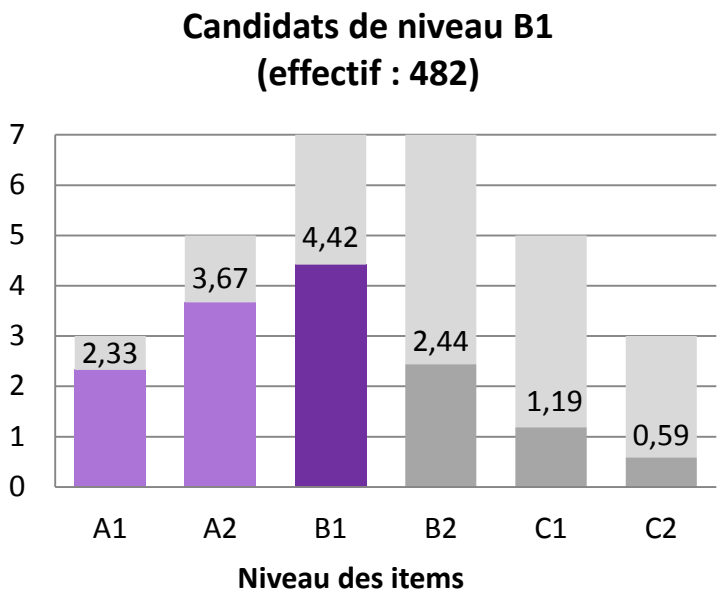
- Version TCF tout public identique pour tous les candidats

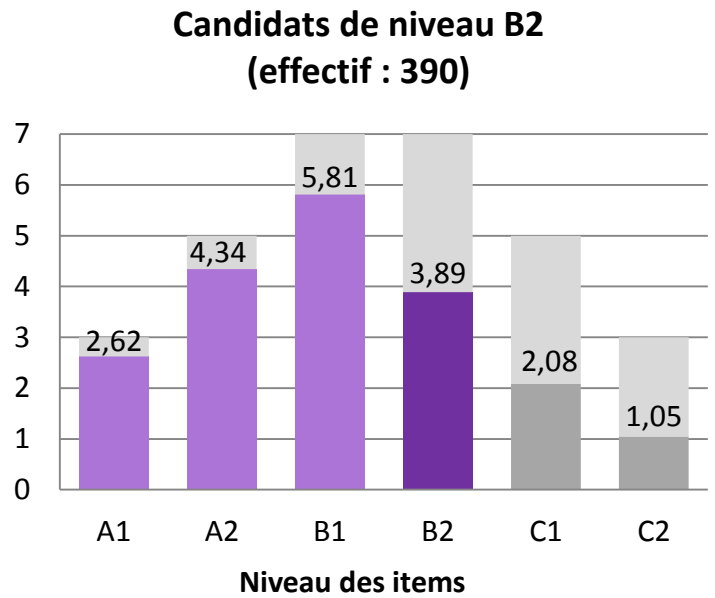
Constats

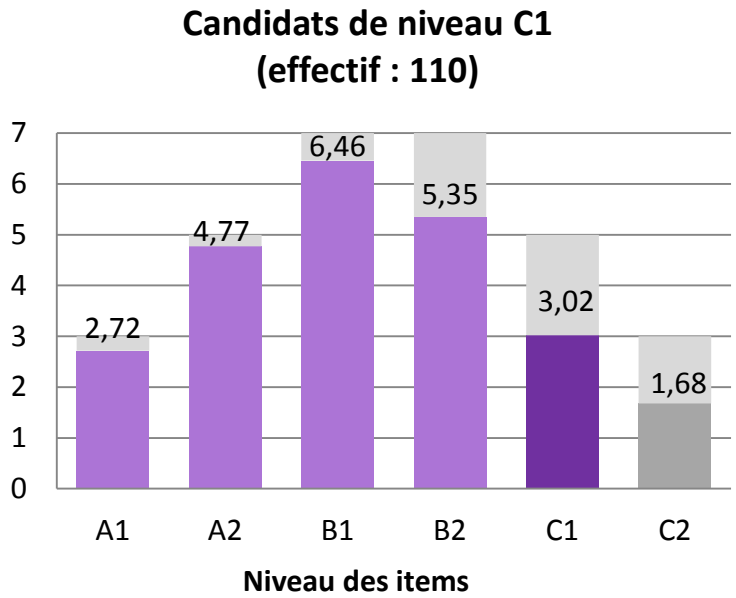
- Nombre moyen d'items réussis en fonction du niveau des candidats et du niveau de difficulté des items.

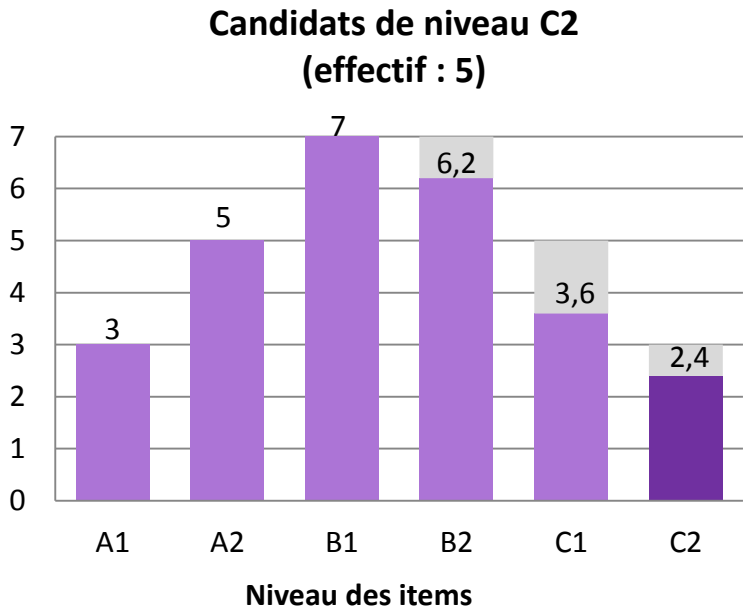





Candidats de niveau B1
(effectif : 482)
7
6
5
4
3
2
1
0
2,33
3,67
4,42
2,44
1,19
0,59
A1
A2
B1
B2
C1
C2
Niveau des items


Candidats de niveau B2
(effectif : 390)
7
6
5
4
3
2
1
0
2,62
4,34
5,81
3,89
2,08
1,05
A1
A2
B1
B2
C1
C2
Niveau des items


Candidats de niveau C1
(effectif : 110)
7
6
5
4
3
2
1
0
2,72
4,77
6,46
5,35
3,02
1,68
A1
A2
B1
B2
C1
C2
Niveau des items


Candidats de niveau C2
(effectif : 5)
7
6
5
4
3
2
1
0
3
5
7
6,2
3,6
2,4
A1
A2
B1
B2
C1
C2
Niveau des items

Méthodes de calcul du score

- Un score pour chaque compétence (CO, MSL, CE)

- 3 méthodes de calcul :
  - Condition 1 témoin (Méthode M1)

  mauvaises réponses non pénalisées = méthode actuelle

  - Condition 2 (Méthode M2)

  pénalisation des mauvaises réponses

  - Condition 3 (Méthode M3)

  critère proportionnel de réussite d'items par niveau

Schéma du calcul du score en condition témoin

Exemple de la compréhension orale du TCF

Version	
A1	Item 1 Item 2 …
…	…
C2	… Item 29 Item 30

Schéma du calcul du score en condition témoin

Exemple de la compréhension orale du TCF

Version	
A1	Item 1 Item 2 …
…	…
C2	… Item 29 Item 30

Mesure de difficulté	
Item 1	-3
Item 2	-2,8
…	…
Item 29	
Item 30	3

Schéma du calcul du score en condition témoin

Exemple de la compréhension orale du TCF

Version	
A1	Item 1 Item 2 …
…	…
C2	… Item 29 Item 30

Mesure de difficulté	
Item 1	-3
Item 2	-2,8
…	…
Item 29	
Item 30	3

Mesure de compétence	
1	-4
2	-2,8
…	…
29	3,80
30	

Nombre d'items réussis

ciep
AU SERVICE DU FRANÇAIS ET DE L'ÉDUCATION DANS LE MONDE

Schéma du calcul du score en condition témoin

Exemple de la compréhension orale du TCF

Version	
A1	Item 1 Item 2 …
…	…
C2	… Item 29 Item 30

Mesure de difficulté	
Item 1	-3
Item 2	-2,8
…	…
Item 29	
Item 30	3

Mesure de compétence	
1	-4
2	-2,8
…	…
29	3,80
30	

Score	
0	
1	100
…	…
30	699

Nombre d'items réussis

ciep
AU SERVICE DU FRANÇAIS ET DE L'ÉDUCATION DANS LE MONDE

Les différentes méthodes de calcul du score

Méthode M1 : mauvaises réponses non pénalisées

	Succès	Echec	Non répondu
M1	1	0	0

Méthode M2 : pénalisation des mauvaises réponses

	Succès	Echec	Non répondu
M2.1	1	-1	0
M2.2	1	-0,5	0
M2.3	1	-0,25	0

Méthode M3 : critère proportionnel de réussite d'items par niveau

	Proportion d'items réussis	
	Niveaux inférieurs	Niveau obtenu
M3.1	Au moins **2/3**	Au moins **2/3**
M3.2	Au moins **2/3**	Au moins **1/3**
M3.3	Au moins **2/3**	Au moins **1 item**

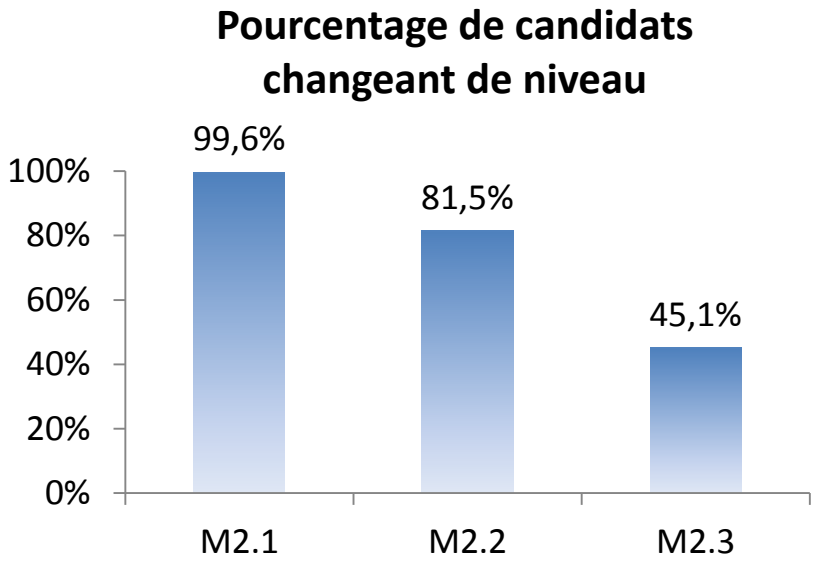
ciep
AU SERVICE DU FRANÇAIS ET DE L'ÉDUCATION DANS LE MONDE

Méthode M2 : pénalisation des mauvaises réponses

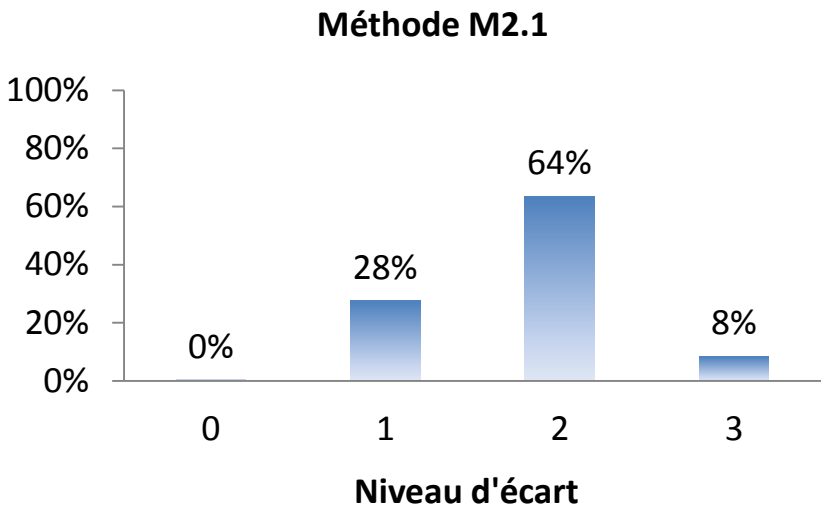
Statistiques sur le score brut obtenu en compréhension orale

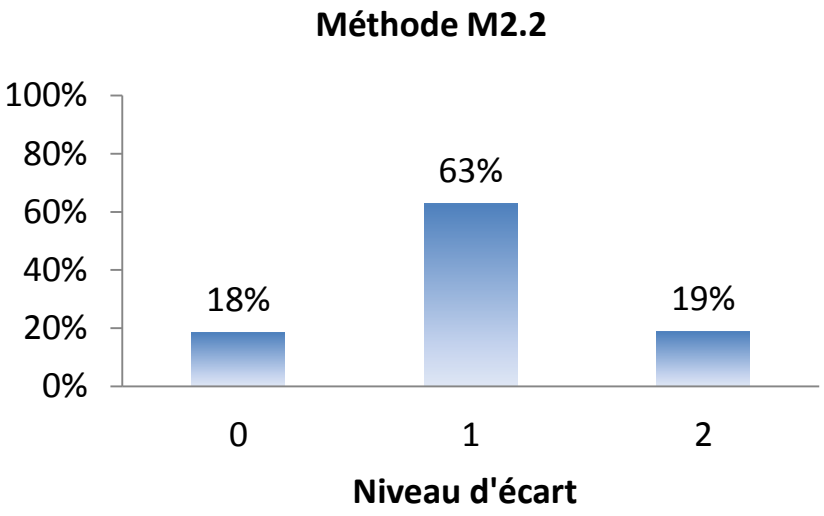
	Succès	Echec	Non répondu
M1	1	0	0
M2.1	1	-1	0
M2.2	1	-0,5	0
M2.3	1	-0,25	0

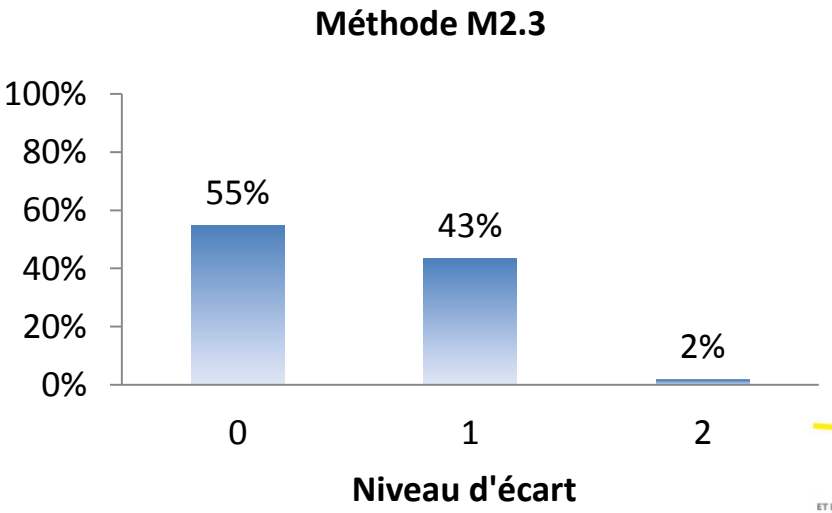
	Moyenne	Ecart-type	Min	Max
M1	17,34	4,09	5	28
M2.1	5	8,10	-19	26
M2.2	11,16	6,09	-7	27
M2.3	14,25	5,09	-1	27,5



Pourcentage de candidats en fonction de l'écart de niveau observé entre la méthode M1 et la méthode M2



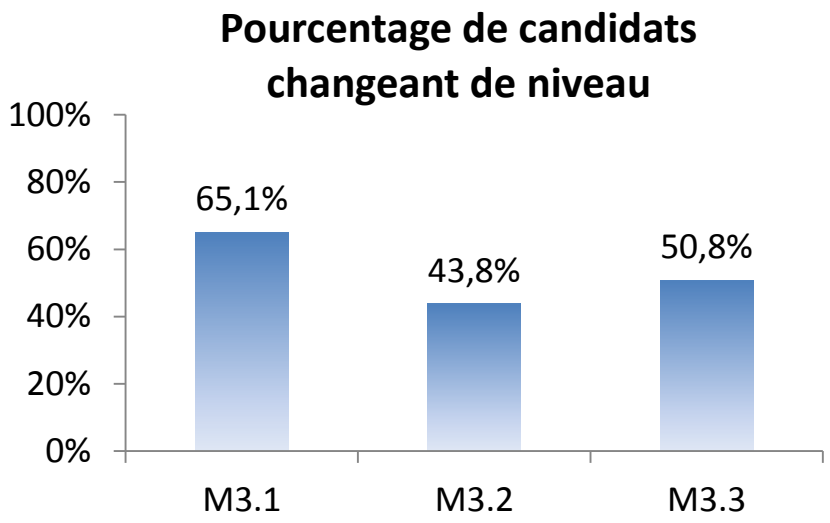




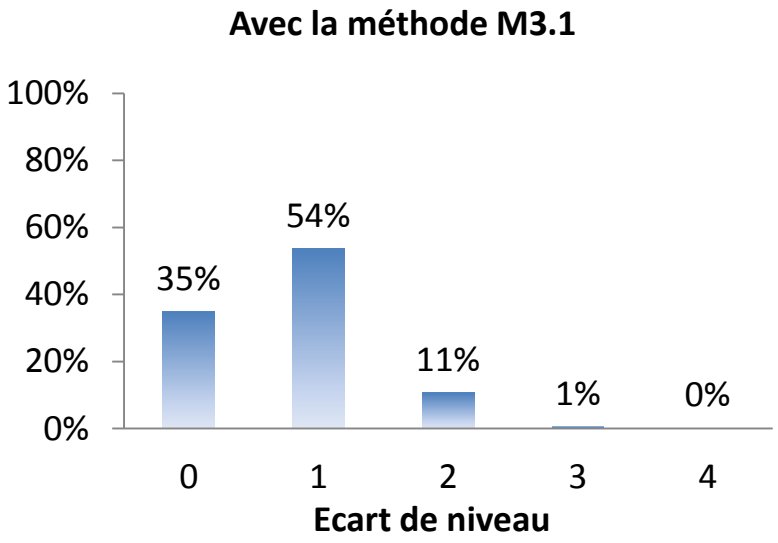
Méthode M3 : critère proportionnel de réussite d'items par niveau

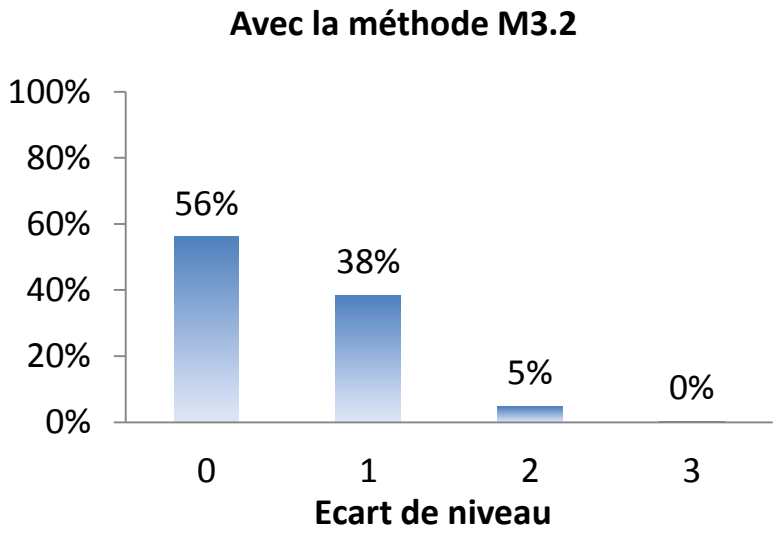
	Succès	Echec	Non répondu
M1	1	0	0

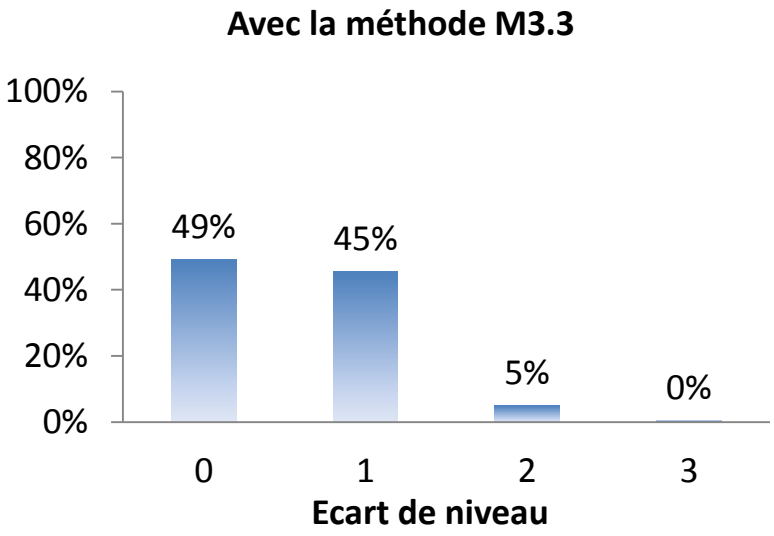
	Proportion d'items réussis	
	Niveaux inférieurs	Niveau obtenu
M3.1	Au moins **2/3**	Au moins **2/3**
M3.2	Au moins **2/3**	Au moins **1/3**
M3.3	Au moins **2/3**	Au moins **1 item**



Pourcentage de candidats en fonction de l'écart de niveau observé entre la méthode M1 et la méthode M3







Pattern de réponses

	Proportion d'items réussis	
	Niveaux inférieurs	Niveau obtenu
M3.1	Au moins **2/3**	Au moins **2/3**

Tous les items de niveau A1 sont réussis

Niveau	A1			A2					B1							B2							C1					C2		
Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Réponses	**V**	**V**	**V**	F	V	F	V	F	V	V	V	V	V	V	V	V	V	V	V	F	V	V	V	F	V	V	F	V	F	V

Moins de 2/3 des items de niveau A2 sont réussis

Pattern de réponses

Pattern de réponses d'un candidat :

- qui a obtenu un **niveau C1** avec la méthode **M1** (23 items réussis)
- qui obtiendrait un **niveau A1** avec la méthode **M3.1** (23 items réussis)

	Proportion d'items réussis	
	Niveaux inférieurs	Niveau obtenu
M3.1	Au moins **2/3**	Au moins **2/3**

Tous les items de niveau A1 sont réussis

Niveau	A1			A2					B1							B2							C1					C2		
Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Réponses	V	V	V	F	V	F	V	F	V	V	V	V	V	V	V	V	V	V	V	F	V	V	V	F	V	V	F	V	F	V

Moins de 2/3 des items de niveau A2 sont réussis

Pattern de réponses de candidats :

- qui ont obtenu un **niveau B2** avec la méthode **M1**
- qui obtiendraient un **niveau A1** avec les **3** méthodes **M3**

	Proportion d'items réussis	
	Niveaux inférieurs	Niveau obtenu
M3.1	Au moins **2/3**	Au moins **2/3**
M3.2	Au moins **2/3**	Au moins **1/3**
M3.3	Au moins **2/3**	Au moins **1 item**

Niveau	A1			A2					B1							B2							C1					C2		
ITEM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Ind1	F	F	V	V	V	F	F	V	V	V	V	V	V	F	V	F	V	F	V	F	V	F	V	V	F	F	V	V	V	F
Ind2	F	V	F	V	V	V	V	V	V	V	V	V	V	V	V	V	F	V	F	F	V	F	V	F	F		V	F	F	F
Ind3	F	V	F	V	F	V	V	V	V	V	V	V	V	V	V	V	F	F	F	F	V	V	V	F	V	F	F	V	V	F

Moins de 2/3 des items de niveau A1 sont réussis

Pattern de réponses de candidats :

- qui ont obtenu un **niveau B1** avec la méthode **M1**
- mais qui obtiendraient un **niveau A1** avec les **3** méthodes **M2**

	Succès	Echec	Non répondu
M2.1	1	-1	0
M2.2	1	-0,5	0
M2.3	1	-0,25	0

Niveau	A1			A2					B1							B2							C1					C2		
ITEM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Ind1	F	V	V	V	V	V	V	F	V	V	V	F	F	V	F	F	F	F	F	F	F	F	F	F	F	F	F	F	V	F
Ind2	F	V	V	V	V	V	F	V	V	F	V	V	F	F	F	F	F	F	F	V	F	V	F	F	F	F	F	F	F	F
Ind3	F	V	V	F	F	V	V	V	V	V	V	F	V	V	F	F	F	F	F	F	F	V	F	F	F	F	F	F	F	F
Ind4	F	V	V	V	V	V	F	F	F	F	F	V	F	F	F	V	F	V	F	F	V	V	F	F	F	F	F	F	V	F

Ils ont tous 11 bonnes réponses et 19 mauvaises réponses.

Conclusion

- Méthode M2
  - → sous-estimation du niveau par rapport à la méthode M1
  - → 1 ou 2 niveaux d'écart (en moins)
  - → plus la pénalité est forte plus l'écart de niveau observé est important

Conclusion

- Méthode M2
  - → sous-estimation du niveau par rapport à la méthode M1
  - → 1 ou 2 niveaux d'écart (en moins)
  - → plus la pénalité est forte plus l'écart de niveau observé est important

- Méthode M3
  - → écart des niveaux dans les deux sens
  - → écart des niveaux moins important qu'avec M2 : un nombre plus important de candidats conserve le même niveau qu'avec la méthode M1
  - → problème d'attribution du niveau pour les candidats qui échouent en début de test

Conclusion

- Méthode M2
  - → sous-estimation du niveau par rapport à la méthode M1
  - → 1 ou 2 niveaux d'écart (en moins)
  - → plus la pénalité est forte plus l'écart de niveau observé est important

- Méthode M3
  - → écart des niveaux dans les deux sens
  - → écart des niveaux moins important qu'avec M2 : un nombre plus important de candidats conserve le même niveau qu'avec la méthode M1
  - → problème d'attribution du niveau pour les candidats qui échouent en début de test

- **Conclusion générale**
  - Sans consigne particulière pour les candidats → la méthode M1 (calcul des habiletés selon Rasch + « fine-equating ») apparait être la seule à pouvoir prendre compte l'ensemble des patterns de réponses

ciep

AU SERVICE
DU FRANÇAIS
ET DE L'ÉDUCATION
DANS LE MONDE

Pour en savoir plus

www.ciep.fr

CIEP sur Facebook

Nous contacter

www.ciep.fr/contact/

Nous rencontrer

1, avenue Léon-Journault, 92318 Sèvres – France