dans le document F2, n’a pas été déclenchée : la grille complète a été administrée en un seul envoi.

VN-Ch07-7 : Paramètres des analyses de sensibilité, cohérence interne et indices des items

Complément du chapitre 7, section 7.7.1

7.6.2. Absence de pré-test et cohérence interne

La cohérence interne des échelles est estimée par le coefficient alpha de Cronbach (1951), calculé par sous-échelle et non pour l’instrument entier, un alpha global sur un questionnaire multidimensionnel n’ayant guère de sens. La valeur de 0,70 sert de repère usuel de lecture. Faute de pré-test, les coefficients ont été calculés après le recueil, sur les effectifs complets, et ils sont rapportés au chapitre 9.

7.6.4. Qualité des items du test de connaissances

La quatrième question est propre au test d’acquisition. Après la passation en T0, chaque item fait l’objet de deux indices classiques. L’indice de difficulté rapporte le nombre de réponses correctes à l’effectif : un item réussi par la quasi-totalité ou échoué par la quasi-totalité n’apporte aucune information discriminante. L’indice de discrimination, estimé par la corrélation point-bisériale corrigée, mesure la capacité de l’item à séparer les élèves qui réussissent l’ensemble du test de ceux qui y échouent : un item que les meilleurs réussissent moins souvent que les plus faibles est un item défectueux.

7.7.1. Traitement des données quantitatives

Des analyses de sensibilité accompagnent ces traitements. Elles ont été définies après le recueil, au moment des analyses, et elles ne modifient aucun critère : elles disent la solidité des résultats. Le rééchantillonnage avec remise (Efron et Tibshirani, 1993), 5 000 tirages au même effectif et à graine fixée, éprouve la stabilité des classements et des moyennes ; pour l’accord entre experts, ce sont les critères et non les experts qui sont rééchantillonnés, la duplication d’un expert gonflant artificiellement l’accord. Un test de permutation, sur 10 000 permutations des étiquettes de classe, éprouve l’écart entre les deux classes de l’étude 2 sans hypothèse de distribution. Une analyse multivers (Steegen et al., 2016) rapporte ensemble tous les choix défendables de score et de modèle. La sensibilité à l’effet de grappe recalcule l’erreur type de

l’écart entre classes en la multipliant par l’effet de plan, pour des corrélations intraclasses de 0 à 0,30, puisqu’une classe par condition ne permet pas d’estimer cette corrélation.

La puissance, estimée par simulation, et l’effet minimal détectable servent à lire les résultats nuls : une absence de différence n’est jamais rapportée sans l’effet que l’effectif permettait de détecter. Les comparaisons multiples sont corrigées par la méthode de Holm (1979), famille de tests par famille de tests. La règle de nettoyage R3, qui écarte les réponses à durée de passation trop courte, est enfin éprouvée à plusieurs seuils, de zéro au quart, au tiers et à la moitié de la durée médiane. Une réserve vaut pour toutes ces analyses : elles mesurent la stabilité face aux fluctuations d’échantillonnage, jamais le biais de recrutement.