



L'utilisation de protocoles individuels expérimentaux et quasi-expérimentaux en psychologie : aspects théoriques et méthodologiques

Christelle Robert

DANS **L'ANNÉE PSYCHOLOGIQUE** 2019/1 (VOL. 119), PAGES 55 À 96
ÉDITIONS **PRESSES UNIVERSITAIRES DE FRANCE**

ISSN 0003-5033

ISBN 9782130821045

DOI 10.3917/anpsy1.191.0055

Article disponible en ligne à l'adresse

<https://www.cairn.info/revue-l-annee-psychologique-2019-1-page-55.htm>



CAIRN.INFO
MATIÈRES À RÉFLEXION

Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...

Flashez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



Distribution électronique Cairn.info pour Presses Universitaires de France.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

L'utilisation de protocoles individuels expérimentaux et quasi-expérimentaux en psychologie : Aspects théoriques et méthodologiques

Christelle Robert ^{*1}
¹ Université de Bordeaux

RÉSUMÉ

Cet article présente les principes fondamentaux des protocoles individuels expérimentaux et quasi-expérimentaux, tels qu'ils peuvent être utilisés en psychologie, notamment pour évaluer de façon expérimentale l'efficacité d'une intervention auprès d'un seul individu. Après avoir introduit les fondements théoriques des plans de recherche à cas uniques, les caractéristiques méthodologiques des protocoles quasi-expérimentaux et expérimentaux sont décrites. Les particularités de chaque plan expérimental sont détaillées et illustrées par une étude. Enfin, les différentes méthodes d'évaluation des données issues de protocoles individuels sont présentées et discutées.

Mots-clés : protocoles individuels expérimentaux ; plans de recherche à cas unique ; méthodes d'analyse des données issues de protocoles individuels expérimentaux.

The use of experimental and quasi-experimental single-case designs in psychology: Theoretical and methodological aspects

ABSTRACT

This paper presents the basic principles of experimental and quasi-experimental single-case designs, as they can be used in psychology, in particular to evaluate experimentally the effectiveness of an intervention with a single individual. After reviewing the theoretical foundations of single-case research designs, the methodological requirements of quasi-experimental and experimental single-case designs are described. The primary features of each

* Correspondance : Christelle Robert, université de Bordeaux, Laboratoire de psychologie EA 4139, 3 place de la Victoire, F-33076 Bordeaux, France. Tel : +33 (0)5 57 57 18 71. Fax: +33 (0)5 57 57 19 77. E-Mail : Christelle.Robert@u-bordeaux.fr

experimental design are detailed and illustrated by a study. Finally, different methods of evaluating data from single-case designs are presented and discussed.

Keywords: experimental single-case designs; single-case research designs; methods of evaluating data from single-case designs.

Les protocoles individuels expérimentaux (appelés aussi *single-case design*, *single-subject design*, *single-system design*, *n=1 design*) présentent la particularité de permettre la conduite d'une recherche utilisant la démarche expérimentale auprès d'une entité individuelle (i.e., un individu, une famille, une classe,...). Du fait de leur flexibilité qui les rend facilement intégrables dans des situations naturelles (voir Nock, Goldman, Wang & Albano, 2004), ces plans sont particulièrement utilisés dans le cadre de recherches appliquées pour démontrer l'existence de relations causales entre des phénomènes dits cliniques (e.g., examiner l'efficacité d'une intervention sur le comportement du patient). L'intérêt croissant pour ces plans peut être en partie expliqué par le mouvement de pratique fondée sur la preuve (« *evidence-based practice* ») qui encourage les cliniciens à fonder leurs pratiques sur des données scientifiques (e.g., Flay et al., 2005 ; Gottfredson et al., 2015 ; Horner & Kratochwill, 2012). Ces procédures restent toutefois largement moins utilisées que les plans à groupes (voir Smith, 2012), certainement à cause d'une méconnaissance de ses principes de base et de la façon d'exploiter les données obtenues (voir Horner, Swaminathan, Sugai, & Smolkowski, 2012 ; Juhel, 2008).

L'objectif de cet article est de présenter les principes fondamentaux de cette démarche expérimentale très utile en psychologie tant pour la recherche fondamentale que pour la pratique clinique. Après avoir présenté les fondements théoriques et historiques de cette démarche, les différents types de protocoles individuels seront décrits d'un point de vue méthodologique et illustrés par des recherches empiriques. Les avantages et limites seront exposés. Enfin, différentes méthodes (visuelle, statistique, clinique) d'évaluation des données recueillies grâce à cette démarche seront présentées et discutées.

LES PROTOCOLES INDIVIDUELS : FONDEMENTS HISTORIQUES ET THÉORIQUES

L'utilisation des plans de recherche à cas uniques en psychologie n'est pas récente (voir Kazdin, 2011). Au contraire, ces plans ont une longue histoire en psychologie. Au XIX^e siècle, la psychologie a émergé de travaux de philosophie et de physiologie (voir Nicolas, 2016) et depuis ses premiers jours en tant que science, elle s'est appuyée sur des études de cas uniques. Ainsi, la plupart des « pères fondateurs » de la psychologie (e.g., Ebbinghaus, Pavlov, Thorndike, Wundt) ont réalisé des expériences contrôlées sur un seul individu (ou sur de petits effectifs), mettant en évidence des phénomènes maintenant devenus classiques dans différents domaines de la psychologie (e.g., perception, apprentissage, mémoire, conditionnement). Par exemple, Ebbinghaus (1913) a réalisé un ensemble de recherches sur l'apprentissage et la mémoire sur une période de 5 ans en utilisant un seul participant : lui-même ! Ces travaux ont notamment permis de mettre en évidence des phénomènes tels que les effets sériels (i.e., effet de primauté et de récence) ou les effets de répétition (i.e., sur-apprentissage, ré-apprentissage, distribution de l'apprentissage) sur la mémoire. L'utilisation des cas uniques a aussi été une méthode privilégiée en psychologie clinique, notamment par Freud et ses collaborateurs dans le cadre d'études de cas dites « naturalistes » (e.g., Freud & Breuer, 1895). Si les méthodes de cas uniques ont permis de réaliser des avancées significatives en psychologie, on peut s'interroger sur les raisons qui ont causé leur abandon par la suite (Stapleton & Hawkins, 2015). On peut supposer que le développement croissant de dispositifs sophistiqués permettant le recueil et l'analyse d'ensembles importants de données quantitatives explique en partie ce phénomène. Partant du principe que plus les données sont nombreuses, plus les résultats obtenus sont valides, les chercheurs ont naturellement privilégié une démarche expérimentale impliquant un effectif élevé de participants et des évaluations statistiques plutôt que des études de cas uniques. Ils ont cependant oublié que les études de cas uniques (expérimentales ou quasi-expérimentales) reposent également sur une démarche expérimentale et que les données obtenues peuvent également être analysées en utilisant des procédures statistiques (voir Nock, Michel, & Photos, 2007).

Dans cet article, les trois protocoles de recherche à cas unique les plus utilisés en psychologie seront présentés : (1) les études de cas « classiques », (2) les protocoles individuels quasi-expérimentaux et (3) les protocoles expérimentaux. Ces protocoles ont pour point commun de se

centrer sur l'étude d'une seule entité (i.e., un seul participant ou un tout petit groupe). Cependant, ils sont caractérisés par un niveau de rigueur scientifique croissant, ce qui par conséquent, permet de les distinguer par la force des inférences (i.e., la relation de cause à effet inférée) que l'on peut faire en les utilisant.

LES ÉTUDES DE CAS

Les études de cas sont des approches souvent utilisées pour observer un individu qui interagit avec des variables d'intérêt (e.g., événements de vie, intervention psychologique). Il y a plusieurs types d'études de cas, mais elles partagent toutes certaines caractéristiques fondamentales (voir Kazdin, 2011). Tout d'abord, les études de cas se centrent sur l'étude d'un individu (i.e., personne, famille, groupe, classe). Le focus de l'étude est sur une unité unique, ce qui permet au chercheur de repérer la complexité liée à l'histoire et aux caractéristiques spécifiques de cet individu. Les données recueillies dans les études de cas sont détaillées, qualitatives et pleines d'anecdotes ; elles sont rarement systématiques ou mesurées de façon quantitative. Enfin, les données sont généralement recueillies de façon rétrospective sans qu'aucun contrôle expérimental ne puisse être appliqué.

Bien que basées sur l'observation, ces méthodes offrent des avantages non négligeables par rapport aux études classiques réalisées sur des groupes. Par exemple, ces études ne nécessitent pas de condition ou de groupe contrôle, ce qui les rend facilement intégrables dans le travail clinique habituel sans gêner le déroulement du traitement. Les études de cas offrent ainsi au clinicien non seulement un outil de recherche pertinent mais aussi une source de données directement applicables à sa pratique clinique. Les études de cas sont aussi très intéressantes pour l'étude de phénomènes rares. L'étude d'Oliver Sacks concernant le Dr. P qui percevait la tête de sa femme comme son propre chapeau en est une bonne illustration (Sacks, 1992). Lorsque les pathologies sont sévères et sont si rares que cela rend impossible le recrutement d'un groupe de participants, les études de cas peuvent fournir des informations très intéressantes sur les phénomènes d'intérêt. De même, l'absence de contraintes ou de cadre expérimental lié aux techniques de mesure offre une opportunité d'observer les comportements tels qu'ils interviennent naturellement, ce qui peut être utile pour générer des hypothèses qui seront ensuite

investiguées en utilisant des méthodes expérimentales. Les études de cas sont aussi très utiles pour le développement et l'évaluation de techniques de traitement nouvelles. Développer et tester des techniques de traitement est coûteux. Dès lors, les études de cas peuvent permettre de récolter des données sur la faisabilité d'une étude plus conséquente (e.g., Moras, Telfer, & Barlow, 1993).

Bien que les études de cas présentent certains avantages, elles manquent de rigueur méthodologique pour permettre de tirer des inférences valides sur les relations entre les variables (Kazdin, 1981). En effet, les études de cas n'intègrent généralement pas d'aspects expérimentaux tels qu'un recueil et une analyse de données systématiques, une manipulation de variable indépendante, ou une réplication des effets. Par conséquent, la validité interne des études utilisant ce type de plan est limitée. Comme nous allons le voir, la validité peut être renforcée en intégrant des éléments expérimentaux au protocole individuel.

LES PROTOCOLES INDIVIDUELS QUASI-EXPÉRIMENTAUX

D'un côté, se trouvent les études de cas qui rendent difficile le contrôle de la situation expérimentale, en partie à cause de leur caractère rétrospectif. D'un autre côté, se trouvent les études expérimentales « pures » qui sont caractérisées par une mesure systématique et une manipulation des variables de façon à démontrer que, toutes choses égales par ailleurs, la variation systématique de la variable indépendante est associée à une modification de la variable dépendante. La possibilité d'inférer une relation causale entre les deux types de variables est ici maximale. Il existe aussi une situation intermédiaire entre l'étude de cas et le protocole expérimental. Dans les situations où le niveau de contrôle et de manipulation requis par un plan purement expérimental n'est pas possible, un plan dit « quasi-expérimental », dans lequel certains éléments d'un plan expérimental sont intégrés, peut être une alternative intéressante. L'ajout d'éléments expérimentaux permet en effet au chercheur d'augmenter la validité interne de son étude par rapport à une simple étude de cas.

Caractéristiques méthodologiques des plans quasi-expérimentaux

Il est à noter que l'ensemble des caractéristiques des plans quasi-expérimentaux détaillées ci-dessous fait partie intégrante des plans expérimentaux.

La répétition des évaluations au travers des différentes phases de l'étude

Le pré-requis méthodologique le plus important pour un protocole individuel quasi-expérimental (et expérimental) est l'évaluation répétée de la variable dépendante au travers des différentes phases de l'étude. La nécessité de répéter l'évaluation du comportement de l'individu est liée au fait que le protocole individuel permet de mettre en évidence des patrons de modifications intra-individuelles. En effet, plutôt que de révéler des différences de moyennes entre des conditions (comme c'est le cas dans la plupart des plans inter-groupes), le protocole individuel permet d'identifier au cours du temps un changement de comportement de l'individu placé dans des conditions expérimentales différentes.

Les évaluations de la variable dépendante doivent être réalisées le plus tôt possible lors de la mise en place de l'étude afin de pouvoir établir une ligne de base, ou un niveau de réponse « pré-intervention ». De plus, les mesures doivent être suffisamment fréquentes pour permettre de mettre en évidence tout changement susceptible d'intervenir pendant l'intervention. Le moment précis choisi de même que la fréquence des évaluations varient selon le phénomène étudié, selon la méthode d'évaluation utilisée et selon les hypothèses du chercheur concernant la rapidité du potentiel changement. Ainsi, l'évaluation peut consister en une mesure constante sur une seule session, un échantillon de comportements mesurés sur des intervalles de temps précis, une mesure quotidienne, etc (voir Nock & Kurtz, 2005). Enfin, il est important d'utiliser des méthodes et outils d'évaluation valides et fidèles pour s'assurer que l'on mesure bien le comportement d'intérêt. Il est également tout aussi important d'utiliser la même procédure d'évaluation au travers des différentes évaluations afin d'éviter que les changements comportementaux observés soient dus à des changements de méthode.

Dans son choix d'utiliser une méthode d'évaluation particulière, l'expérimentateur et/ou le praticien doit trouver le juste équilibre entre

l'utilisation de méthodes fidèles, valides et pertinentes et la faisabilité par rapport au contexte clinique et aux moyens financiers et humains dont il dispose. L'expérimentateur doit également faire attention à restreindre les évaluations réalisées aux seules mesures nécessaires pour sa recherche, au risque de diminuer l'adhérence du participant au protocole. Pour enrichir ses évaluations, il peut tout à fait intégrer des informations externes et des données déjà collectées (e.g., journaux quotidiens complétés par le patient ou un proche, données issues des membres de la famille, enseignants, éducateurs, personnels soignants,...).

L'utilisation d'une intervention ou d'une manipulation expérimentale spécifique

Une autre caractéristique méthodologique très importante d'un protocole individuel quasi-expérimental (et expérimental) est la mise en place d'une intervention (e.g., traitement, thérapie) ou d'une manipulation expérimentale spécifique. D'un point de vue méthodologique, l'intervention qui est mise en place correspond à la variable indépendante de l'étude. Afin de pouvoir faire des prédictions (i.e., poser des hypothèses) et tirer des conclusions claires et interprétables sur la relation causale entre l'intervention et le comportement produit, il est nécessaire de savoir très précisément en quoi consiste l'intervention mise en place. L'opérationnalisation précise de la variable indépendante est également un élément important pour permettre une réplification de l'étude.

Exemple d'une étude utilisant un protocole individuel quasi-expérimental

Paulus et Norton (2016) ont utilisé un protocole individuel quasi-expérimental visant à évaluer l'efficacité d'une thérapie comportementale pour traiter une phobie spécifique (i.e., émétophobie ou peur de vomir) chez une jeune femme de 23 ans. Une thérapie cognitivo-comportementale basée sur une approche transdiagnostique a été mise au point (voir Paulus & Norton, 2016, pour la description précise du traitement). Ce traitement, comportant différentes étapes clairement spécifiées et réparties sur 5 sessions, a été administré de façon contrôlée tout au long du traitement. Plusieurs mesures ont été relevées de façon hebdomadaire avant chaque session d'intervention. Il s'agissait notamment d'évaluer le degré d'émétophobie (échelle *SPOVI* allant de 0 à 56, Veale et al., 2013),

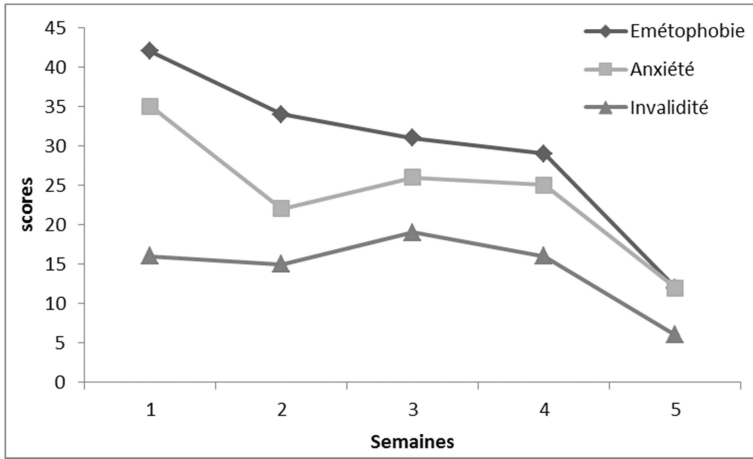


Figure 1. Indices d'émétophobie, d'anxiété et d'invalidité recueillis de façon hebdomadaire avant chaque session de traitement de Jane (voir Paulus & Norton, 2016).

Figure 1. *Emetophobia, anxiety and disability indices collected weekly before each Jane's treatment session (see Paulus & Norton, 2016).*

d'anxiété (échelle *ADDQ-W*, Smith, Paulus, & Norton, 2017) et d'invalidité (échelle *SDS*, Sheehan, 1983). La Figure 1¹ présente les différents indices recueillis auprès de la patiente tout au long du traitement.

De façon générale, les différents indices ont diminué entre la session 1 et la session 5, cette diminution étant plus marquée entre les sessions 4 et 5. Les scores d'émétophobie et d'anxiété, initialement très élevés, se sont retrouvés en dessous du seuil pathologique en fin de protocole. Il est à noter qu'une augmentation des indices d'anxiété et d'invalidité est survenue entre les sessions 2 et 3 en lien avec un évènement stressant que la patiente a vécu durant cette période (i.e., elle a dû quitter précipitamment un dîner à cause de maux de ventres et d'une appréhension de vomir). À la fin de la 5^e séance, la patiente a souhaité arrêter l'intervention dans la mesure où elle n'en ressentait plus la nécessité.

Dans l'étude de Paulus et Norton (2016), l'utilisation d'évaluations répétées et d'une intervention bien spécifiée fournit des informations intéressantes non seulement sur l'intervention utilisée, mais aussi sur la durée et l'évolution du changement de comportement au cours du traitement. Cependant, la principale faiblesse de cette étude est qu'elle n'intègre

¹ Les données des graphiques recomposés qui illustrent les études présentées ont été récupérées en utilisant le logiciel DataThief III (Tummers, 2006).

pas de condition contrôle, c'est-à-dire de condition sans intervention, qui permettrait d'estimer le comportement pathologique habituel du patient. Sans cette condition contrôle, on peut tout à fait imaginer que le changement de comportement observé soit la conséquence d'autres facteurs externes à l'étude. L'utilisation d'un protocole individuel expérimental permet de résoudre ce problème.

LES PROTOCOLES INDIVIDUELS EXPÉRIMENTAUX

En plus des éléments expérimentaux décrits dans la section précédente, les protocoles individuels expérimentaux intègrent également des procédures de contrôle de variables parasites liées à l'étude d'un cas unique, permettant ainsi d'augmenter la validité des effets obtenus.

Caractéristiques méthodologiques des plans expérimentaux

L'existence de plusieurs phases ou conditions expérimentales. Dans les protocoles individuels tout comme dans les plans inter-groupes, l'effet de la variable indépendante est évalué en comparant les performances de la/les condition(s) expérimentale(s) à celles d'une condition contrôle. Une différence fondamentale entre les deux types de plan (groupe vs. individu) est liée au mode de contrôle des caractéristiques des participants sur la performance observée. Dans les plans inter-groupes, les participants sont affectés de façon aléatoire soit dans la/les condition(s) expérimentale(s), soit dans la condition contrôle. La randomisation d'un nombre conséquent de participants est supposée garantir un équilibre naturel des caractéristiques des participants au travers des différentes conditions. Dans les protocoles individuels expérimentaux où il n'y a par définition qu'un seul individu, c'est ce même individu qui est placé dans toutes les conditions (expérimentale(s) et contrôle), maintenant ainsi constantes les caractéristiques individuelles au travers des conditions et éliminant leur éventuelle influence sur le comportement observé. Comme nous le verrons par la suite, le nombre et l'ordre des conditions varient en fonction du type de plan expérimental utilisé. Retenons que l'élément méthodologique fondamental est l'inclusion de phases ou conditions expérimentales multiples pour chaque participant.

Classiquement, la première phase d'un protocole individuel expérimental est une condition de base, ou condition contrôle, qui a pour objectif d'évaluer le comportement habituel de l'individu, c'est-à-dire sans intervention. Pour constituer cette condition de base, il est nécessaire de recueillir plusieurs mesures de la variable dépendante, suffisamment pour s'assurer que le comportement mesuré est stable. Obtenir un patron de données stable dans la condition contrôle est fondamental. Non seulement il renseigne sur le comportement habituel du participant que l'on cherche à modifier, mais aussi il permet de faire des prédictions sur le comportement attendu suite à la mise en place de l'intervention. Il est recommandé d'effectuer au minimum entre 3 à 5 mesures dans la condition de base pour avoir confiance dans le patron observé (voir Ganz & Ayres, sous presse ; Horner, Carr, Halle, McGee, Odom, & Wolery, 2005 ; Krasny-Pacini & Evans, 2018). De façon générale, plus le nombre de mesures recueillies est élevé, plus le chercheur pourra avoir confiance dans le patron de données obtenu. Enfin, il faut également faire attention à ne pas obtenir d'effet plafond (ou d'effet plancher) dans la condition de base sous peine de rendre difficile l'observation de modifications effectives lorsque l'intervention sera mise en place.

Obtenir un patron de données stable ne signifie pas forcément recueillir exactement les mêmes scores au travers des différentes mesures. Différents patrons de données peuvent refléter une performance stable. Tout d'abord, une ligne de performance stable peut être observée lorsque les scores sont relativement similaires, c'est-à-dire lorsqu'il y a très peu de variabilité. Ce patron de données est idéal car il permet de détecter très facilement un changement lorsque l'intervention est mise en place. Ensuite, une tendance linéaire (i.e., les scores augmentent ou diminuent de façon constante au cours du temps) ou curvilinéaire (i.e., le degré de changement s'accélère ou diminue au cours du temps) représente également un patron de données stable. Enfin, un patron de données en cycles peut également être observé. Dans ce cas-là, les scores augmentent et diminuent selon certaines périodes, par exemple selon le moment du mois ou de l'année. Le trouble affectif saisonnier, conséquence d'un manque de lumière qui dérègle la production de mélatonine et de sérotonine, est un bon exemple de modifications cycliques annuelles d'un état d'humeur (voir Melrose, 2015). Malheureusement, on peut aussi être confronté à des situations dans lesquelles aucun patron stable n'émerge. Ce sont des patrons de données qui présentent une forte variabilité sans explication apparente. Avant toute chose, il est important de considérer les raisons de cette variabilité élevée. Notamment, on peut s'interroger sur les qualités psychométriques, notamment la fidélité, de l'instrument de mesure utilisé,

et envisager de changer d'instrument. Il faut également s'assurer que l'instrument soit utilisé correctement par le patient ou par toute autre personne qui est supposée l'utiliser (e.g., membre de la famille, enseignant, personnel soignant). Enfin, il s'agira aussi de vérifier qu'aucun événement particulier ne soit survenu dans la vie du patient, pouvant rendre compte d'un comportement instable.

L'efficacité d'une intervention est testée en évaluant la présence d'un changement de la variable dépendante suite à l'introduction de changements systématiques dans une ou plusieurs conditions. Si toutes choses étant égales par ailleurs et que la seule variation est l'intervention, alors tout changement de la variable dépendante peut être imputé à cette intervention. Bien sûr, il est rarement possible de contrôler toutes les variables parasites externes, et il est donc toujours possible que des facteurs autres que l'intervention aient causé cette modification. Ce genre de menaces à la validité interne de l'étude peut être éliminé en cherchant à répliquer les effets expérimentaux obtenus.

La réplication des effets expérimentaux

Répliquer un effet est une condition nécessaire pour garantir sa validité. Dans le cadre des protocoles individuels expérimentaux, les effets expérimentaux peuvent être répliqués de différentes façons, soit au sein d'une même étude (e.g., au travers du même comportement de plusieurs participants ou au travers de plusieurs comportements d'un même participant), soit au travers de plusieurs études différentes. Nous détaillerons par la suite pour chaque type de plan expérimental les procédures qui permettent la réplication des effets.

Les normes de représentation graphiques des données issues de protocoles individuels expérimentaux

Par convention, les données recueillies à l'aide de protocoles individuels expérimentaux sont représentées sous forme d'un graphique (voir Figure 2). Cette représentation graphique facilite l'évaluation visuelle de l'efficacité de l'intervention. Au sein du graphique, l'axe des ordonnées représente les scores de la variable dépendante et l'axe des abscisses représente les unités de temps (e.g., heures, jours, mois, sessions).

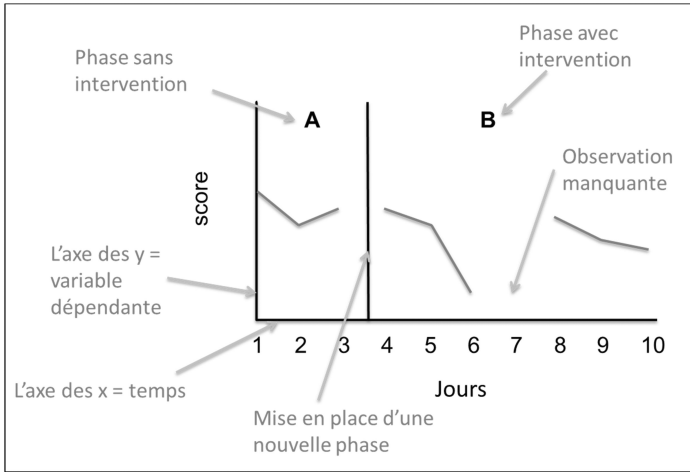


Figure 2. Normes de représentation graphique des données issues de protocoles individuels expérimentaux.

Figure 2. Standards for graphical representation of data from single-case designs.

Les observations manquantes sont représentées par des ruptures de lignes. Les phases ou conditions correspondent à des périodes temporelles distinctes durant lesquelles une manipulation est introduite ou non. Ainsi, la phase de non-intervention (ou ligne de base ou condition contrôle) est représentée par la lettre A, et les phases d'interventions sont représentées par d'autres lettres (B, ou C dans le cas d'une deuxième intervention). Les phases sont séparées par des lignes verticales et les courbes correspondant à la performance de l'individu sont disjointes, indiquant ainsi une rupture entre les conditions des phases adjacentes.

Les différents protocoles individuels expérimentaux

Il existe plusieurs types de protocoles individuels expérimentaux. De plus, certains de ces plans peuvent être combinés, ce qui augmente encore le nombre de plans possible. L'ensemble de ces protocoles respectent les pré-requis méthodologiques détaillés dans la section précédente (i.e., la mise en place d'évaluations répétées, l'utilisation d'une ou de plusieurs interventions bien spécifiées, et l'inclusion de différentes phases). Ils diffèrent dans le nombre et l'ordre des différentes phases et dans la façon de démontrer le lien causal entre l'intervention et le changement de comportement.

Les plans avec retrait (ABA ou ABAB)

Les plans avec retrait, communément appelés plans ABA ou ABAB, impliquent au minimum trois phases. Lors de la première phase (A1), le comportement contrôle (i.e., pré-intervention) du participant est évalué. Puis, l'intervention est mise en place lors d'une seconde phase expérimentale (B1). Enfin, l'intervention est retirée lors de la troisième phase (A2), ce qui signifie qu'il y a un retour à la phase contrôle (A2). On s'attend ici à ce que le comportement du participant revienne à son niveau de base initial.

Les mesures collectées lors de la première phase (A1) servent de condition contrôle et sont nécessaires pour, d'une part, établir le comportement naturel de l'individu (en l'absence d'intervention), et d'autre part, pour démontrer que tout changement comportemental n'est pas lié à des facteurs externes tels que le contact avec le thérapeute ou la mise en place de procédures d'évaluation. La démonstration d'un changement stable de comportement lorsque, et uniquement lorsque, la phase expérimentale débute suggère que c'est bien l'intervention qui est responsable de ce changement de comportement. Il subsiste toutefois toujours la possibilité pour que d'autres facteurs externes aient causé ce changement. On peut citer par exemple (voir Cooper, Heron & Heward, 2007 ; Gast, 2010 ; Kratochwill et al., 2010 pour des discussions des principales variables parasites) les facteurs historiques (i.e., un événement extérieur à l'individu), maturationnels (i.e., un changement physiologique ou psychologique intervenant lors de l'étude) ou la régression vers la moyenne (i.e., les scores extrêmes, très faibles ou très élevés, obtenus lors d'une évaluation initiale ont tendance à se rapprocher de la moyenne lors des évaluations suivantes). Dans les plans avec retrait, ces menaces à la validité interne sont évitées par le retour à la condition de base. Si le comportement du participant revient ou approche son comportement initial de base lorsque l'intervention est retirée (phase A2), le lien causal entre l'intervention et le changement de comportement est renforcé. Après un plan de type ABA, il est habituel de ré-appliquer l'intervention (ce qui conduit à un plan ABAB) de façon non seulement à tenter de répliquer l'effet de l'intervention, mais aussi pour des raisons cliniques et éthiques évidentes.

Xin, Sheppard et Brown (2017) ont évalué l'efficacité d'une intervention pour réduire les comportements scolaires inadaptés (i.e., faire du bruit, sortir de sa chaise, jouer avec des objets,...) de 4 enfants présentant un trouble du spectre autistique. L'intervention consistait à mettre à disposition des élèves une tablette numérique avec une application pour les aider à accomplir des routines quotidiennes et à comprendre et contrôler

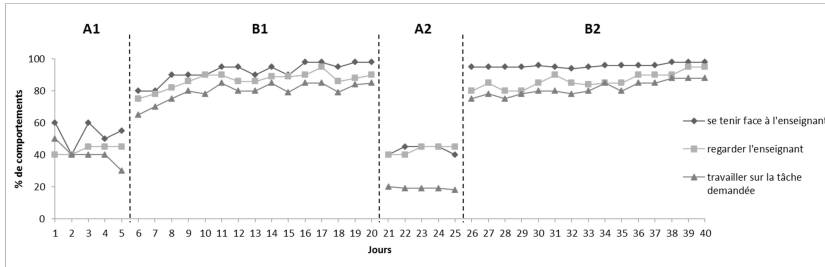


Figure 3. Plan ABAB utilisé pour tester l'efficacité d'une intervention sur les comportements scolaires chez un enfant présentant un trouble du spectre de l'autisme (voir Xin, Sheppard, & Brown, 2017).

Figure 3. ABAB design used to test the effectiveness of an intervention on school behavior in a child with autism spectrum disorder (see Xin, Sheppard, & Brown, 2017).

leurs émotions associées à ces routines. Les auteurs ont utilisé un plan avec retrait de type ABAB (voir Figure 3). Après avoir évalué sur 5 jours la fréquence (au travers de sessions de 20 mn.) de 3 comportements adaptés pour réaliser une tâche donnée (phase A1), l'intervention a été mise en place durant 15 jours (phase B1), montrant une augmentation de la fréquence des comportements adaptés. Puis, l'intervention a été retirée durant 5 jours (phase A2) et une diminution de la fréquence des comportements adaptés a été observée. Enfin, l'intervention a été ré-introduite (phase B2) et les comportements adaptés ont augmenté à nouveau. En outre, il semble que l'intervention ait également eu un impact positif sur les performances académiques des enfants, leur niveau de vocabulaire s'étant amélioré à la fin du protocole.

Démontrer que la présence *vs.* l'absence d'une intervention change de façon répétée la fréquence d'un comportement fournit une preuve indiscutable de causalité entre ces deux variables. Toutefois et même dans le cas où l'intervention provoque un changement de comportement, il peut être difficile dans certaines situations de faire revenir le comportement à son niveau de base. Par exemple, si le traitement implique le développement de compétences cognitives (e.g., stratégies mnémoniques), le retour au comportement initial va être difficile, les compétences instaurées ne pouvant pas être supprimées après coup. Il y a aussi des situations pour lesquelles le retour au comportement de base n'est pas souhaitable. Par exemple, si une intervention a diminué les risques de suicide chez un patient, il est préférable tant d'un point de vue clinique qu'éthique, de ne pas revenir au comportement initial (voir Rizvi & Nock, 2008). Dans ces cas-là, d'autres plans expérimentaux sans retrait de l'intervention seront privilégiés.

Les plans à lignes de base multiples

Les plans à lignes de base multiples sont des plans de type A-B pour lesquels la condition expérimentale est introduite au cours du temps sur différents comportements, ou différents contextes ou différents participants. La spécificité de ces plans est de mesurer plusieurs lignes de base. Kazdin (2011) recommande au minimum l'inclusion de 3 lignes de base. Une fois les lignes de base établies, une manipulation expérimentale est appliquée sur l'un de ces comportements pour observer si ce comportement cible change alors que les autres comportements restent stables.

Trois types de plans à lignes de base multiples peuvent être utilisés. Tout d'abord, les plans à lignes de base multiples au travers de comportements consistent à ce que la même intervention soit appliquée à des moments différents à des comportements cibles indépendants chez un seul individu. Ensuite, les plans à lignes de base multiples au travers de participants impliquent qu'un traitement particulier soit appliqué à des moments différents à des participants appariés et supposés avoir été exposés à des conditions environnementales identiques. On se centre ici généralement sur un seul comportement cible. Enfin, les plans à lignes de base multiples au travers de contextes consistent à appliquer un traitement particulier à des moments différents à un seul participant placé dans des contextes différents et indépendants (e.g., école, maison). On se centre également généralement sur un seul comportement cible.

L'efficacité de l'intervention est ici démontrée lorsque la modification du comportement ne se produit que lorsque l'intervention est dirigée sur ce comportement (ou le contexte ou le participant) en question, alors que les autres lignes de base (i.e., comportements, contextes, ou participants) restent stables. Comme pour les plans à retrait, les plans à lignes de base multiples débutent par une condition de base qui est suivie par la condition expérimentale. Un changement de comportement à ce moment-là suggère que l'intervention en est la cause. Cependant et comme indiqué précédemment, il faut prendre des précautions pour éviter l'influence de facteurs externes. Plutôt que d'utiliser un retour à la ligne de base, les plans à lignes de base multiples cherchent à répliquer l'effet de l'intervention. Ainsi, l'intervention va être appliquée à différents comportements, différents contextes ou différents participants au fil du temps. Le décalage temporel de l'introduction de l'intervention est l'élément fondamental qui permet d'éviter que des facteurs externes expliquent le changement de comportement observé. L'expérimentateur s'assure de l'efficacité de l'intervention lorsqu'un changement comportemental intervient après sa mise en place alors que les autres comportements (i.e., non traités) restent

stables. Ce type de plan part donc du principe que les comportements ciblés sont indépendants les uns des autres.

Hendricks et al. (2017) ont utilisé un plan à lignes de base multiples au travers de différents participants pour évaluer l'efficacité d'une intervention comportementale auprès de 10 adolescents âgés entre 12 et 18 ans présentant un état de stress post-traumatique. Une ligne de base dont la durée variait de 4 à 8 semaines a tout d'abord été constituée. Les participants étaient affectés de façon aléatoire à une des 5 lignes de base pré-définies (i.e., 4, 5, 6, 7 ou 8 semaines à raison de 2 participants par ligne de base). Puis, l'intervention était mise en place. Elle consistait en une exposition prolongée intensive à l'évènement traumatique et était administrée pendant 5 jours consécutifs, à raison de 3 sessions quotidiennes de 90 min. Lors de la première session, il était demandé au patient de revivre mentalement la scène traumatique et de la raconter à voix haute en donnant le plus de détails possible. La deuxième session consistait en une exposition à l'évènement traumatique par le dessin : le patient devait dessiner la scène en y intégrant tous les détails anxiogènes. Enfin, une exposition *in vivo* était mise en place lors de la troisième session, le patient étant alors confronté à un matériel et/ou à des situations directement liées à l'évènement traumatique. Un suivi à 3 mois a été réalisé sur 4 semaines pour vérifier la stabilité des potentiels changements. Durant l'étude, les parents ont également participé à des séances visant à améliorer la qualité de leur soutien social lors des échanges avec leur enfant concernant l'évènement traumatique. Un indice de sévérité des symptômes de l'état de stress post-traumatique a été relevé (voir Steinberg, Brymer, Decker, & Pynoos, 2004), de même que plusieurs mesures de dépression, d'anxiété généralisée et de symptômes dissociatifs (voir Hendriks et al., 2017, pour une description précise des différentes échelles utilisées). Les résultats, présentés dans la Figure 4, indiquent que l'indice de sévérité des symptômes de l'état de stress post-traumatique a diminué pour l'ensemble des 10 participants quand, et seulement quand, l'intervention a été appliquée au participant. Ce patron de données démontre ainsi que c'est l'intervention, et non des facteurs externes, qui a provoqué le changement de comportement.

Parmi les points forts des plans à lignes de base multiples, on peut citer leur caractère flexible dans le sens où ils sont facilement applicables dans un contexte clinique ainsi que leur objectif de réplication des effets, critère nécessaire pour démontrer un lien causal. Enfin, les plans à lignes de base multiples n'impliquent pas de retrait d'une intervention, ce qui élimine les considérations cliniques et éthiques souvent avancées à l'encontre des plans avec retrait. Une limite potentielle de ces plans est

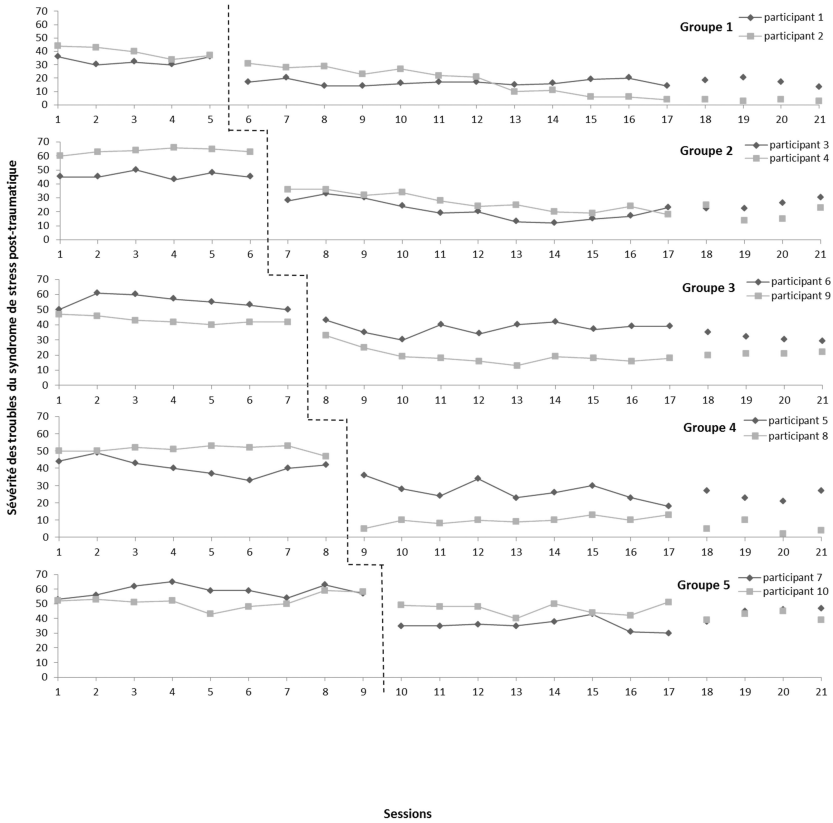


Figure 4. Plan à lignes de base multiples au travers de 5 groupes de 2 patients présentant un état de stress post-traumatique (voir Hendricks et al., 2017).

Figure 4. Multiple baseline design across 5 groups of 2 patients with post-traumatic stress disorder (see Hendricks et al., 2017).

liée à l'indépendance des critères choisis (i.e., comportements, contextes ou participants). En effet, pour démontrer le lien de causalité, le changement comportemental ne doit pas intervenir sur les autres critères avant que l'intervention ne soit dirigée vers eux. Cela nécessite donc de choisir en amont des critères qui soient indépendants les uns des autres. Si les comportements ciblés co-varient (i.e., un changement est observé sur toutes les lignes de base quand l'intervention est appliquée au premier comportement), les effets observés ne peuvent pas être interprétés et il est nécessaire d'utiliser d'autres types de plans.

Les plans à critères changeants

Les plans à critères changeants sont des plans de type A-B qui impliquent un critère de renforcement du comportement attendu. Ces plans débutent par une condition de base qui est suivie par une phase d'intervention dirigée vers un comportement. La particularité ici est de mettre en place un critère pour renforcer ce comportement, critère progressivement modifié au cours du temps pour permettre d'augmenter l'ampleur du changement du comportement. La relation causale entre l'intervention et le changement comportemental est démontrée par le fait que le comportement change uniquement lorsque le critère de renforcement est modifié. Le changement de critère est opéré par le chercheur et intervient généralement une fois que le comportement est stable avec le critère en cours. Habituellement, le changement est opéré dans la même direction au fil du temps (i.e., soit augmenté, soit diminué), mais le chercheur peut aussi inverser pour une étape le sens du changement pour s'assurer que le comportement de l'individu s'accorde avec.

En utilisant un plan à critères changeants, Cook, Rapp, et Schulze (2015) ont évalué l'efficacité d'une intervention visant à maintenir au poignet un bracelet d'alerte médicale chez un jeune garçon autiste âgé de 8 ans. Le patient présentait des troubles du comportement lorsque le bracelet était mis à son poignet. Il criait, demandait d'enlever le bracelet, se roulait par terre, frappait, tirait ou mordillait le bracelet. L'étude, conduite sur 8 semaines, a commencé par une ligne de base durant laquelle l'intervalle de temps entre la mise au poignet du bracelet et le déclenchement des troubles du comportement était mesuré. En moyenne, cet intervalle était de 5 secondes (voir Figure 5). Puis, l'intervention était mise en place. Elle consistait à avertir le patient du positionnement du bracelet et de son retrait après un temps prédéfini, un minuteur étant activé lors de la pause du bracelet. Si aucun trouble du comportement n'apparaissait pendant la durée fixée, le bracelet était retiré et l'essai était considéré comme réussi. Le critère de renforcement était ici la durée prédéfinie du port du bracelet. Ce critère était progressivement augmenté pour accroître en conséquence la durée de maintien du bracelet sans troubles comportementaux. Vingt-deux critères ont été fixés, allant d'une durée de 5 secondes jusqu'à une durée de 7 heures. La Figure 5 présente les résultats pour la ligne de base et les critères de 5 secondes à 2 minutes. Le changement de critère était globalement associé à un changement du comportement qui allait dans le même sens. Un suivi effectué deux ans après la fin du traitement a permis de s'assurer que le patient portait désormais le bracelet de façon continue.

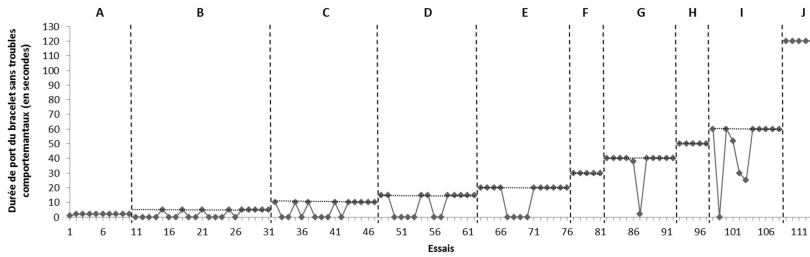


Figure 5. Plan à critères changeants utilisé pour tester l’efficacité d’une intervention sur la durée du port d’un bracelet d’alerte médicale sans troubles du comportement (voir Cook et al., 2015).

Figure 5. Changing criterion design used to test the effectiveness of an intervention on the duration of the wearing of a medical alert bracelet without behavioral disorders (see Cook et al., 2015).

L’utilisation d’un plan à critères changeants présente plusieurs avantages. Notamment, dans le cas où un changement net de comportement n’est pas immédiat suite à la mise en place de l’intervention, le plan à critères changeants est une option intéressante car il permet d’intervenir graduellement et de façon systématique pour modifier le comportement ciblé. De plus, dans la mesure où les plans à critères changeants ne nécessitent pas de retrait du traitement (comme c’est le cas pour les plans à retrait) et ne se focalisent pas sur un traitement figé (comme c’est le cas pour les plans à lignes de base multiples), ce sont ceux qui impliquent le moins de déviations par rapport à une pratique clinique normale, et qui peuvent être facilement intégrés dans ce cadre.

Une limite potentielle à ces plans, c’est qu’il peut être difficile de les utiliser pour démontrer une relation causale entre l’intervention et le changement de comportement si le comportement cible ne suit pas le critère de renforcement. Prenons par exemple l’étude de Hodges, Davis, Crandall, Phipps, et Weston (2017) visant à tester l’efficacité d’une intervention pour augmenter la consommation d’aliments variés chez deux enfants qui refusaient de s’alimenter. Trois critères de renforcement étaient utilisés (1 : porter l’aliment à ses lèvres, 2 : mettre l’aliment dans sa bouche, et 3 : mâcher l’aliment et l’avaler). Les résultats ont indiqué que les comportements des enfants suivaient les critères fixés par l’expérimentateur, démontrant l’efficacité de l’intervention. Cependant, si durant la 1^{re} phase d’intervention où le critère de renforcement était de 1 (i.e., porter l’aliment à ses lèvres), le patient avait produit immédiatement et de façon stable un comportement alimentaire qui s’apparentait au critère 3 (i.e., mâcher l’aliment et l’avaler), il aurait été très difficile de conclure

que l'intervention était la cause de ce changement net et rapide de comportement. Bien que de telles modifications puissent être considérées comme un réel succès d'un point de vue clinique, elles ne permettent pas de comprendre les raisons de ce changement.

Les plans à traitements multiples

Les plans expérimentaux présentés précédemment se limitent généralement à l'évaluation de l'efficacité d'une seule intervention. Dans certains cas, le chercheur peut être cependant intéressé par la comparaison de deux ou plusieurs interventions afin d'identifier la plus efficace. Il est alors possible d'utiliser des plans à traitements multiples.

Un premier type de plans à traitements multiples consiste à appliquer les différents traitements au même participant au travers de plans de type ABAB ou de plans à lignes de base multiples (e.g., Ingram, Lewis-Palmer, & Sunai, 2005). Dans ce cas, les différents traitements sont introduits dans des phases différentes de sorte qu'un des traitements intervient avant le/les autre(s). Ces plans sont appelés A-B-C-D, dans lesquels B, C et D représentent des interventions différentes (voir Halldorsson & Salkovskis, 2017, pour une utilisation récente de ce type de plan) et peuvent être combinés à des plans avec retrait (e.g., ABCABC).

Les plans à traitements alternés forment un autre ensemble de plans à traitements multiples et présentent la particularité d'alterner de façon rapide deux ou plusieurs traitements différents au sein d'une même phase pour en mesurer les effets sur le comportement cible. Ces plans sont les plus appropriés pour déterminer les interventions les plus efficaces parmi celles qui sont testées (voir Kazdin, 2011). Les traitements sont alternés de façon aléatoire, ce qui ne rend pas nécessaire une mesure de la ligne de base. Il faut au minimum comparer deux interventions pour un même participant. Par exemple, l'administration aléatoire de deux traitements B et C peut produire la séquence suivante : B-C-C-B-C-B-B-C. L'effet de l'intervention est appréhendé en comparant les patrons de données des deux interventions (voir Figure 6). Si les données ne se recouvrent pas, on peut conclure qu'une intervention est plus efficace que l'autre. L'effet différentiel des interventions est établi par la distance entre les différents patrons de données. Ainsi, plus la différence est importante, plus l'effet différentiel est important.

Pelaez, Borroto, et Carrow (2018) ont utilisé un plan à traitements alternés pour évaluer l'effet de l'utilisation d'un renforcement social sur

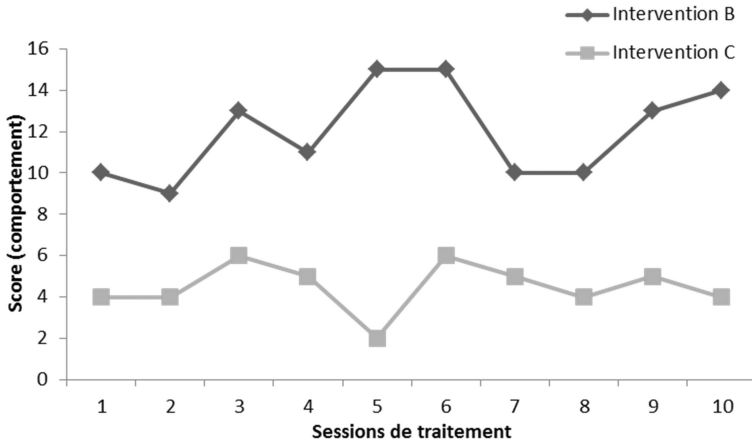


Figure 6. Représentation visuelle de données issues d'un plan à traitements alternés.

Figure 6. Visual representation of data from an alternating treatment design.

les performances de vocalisations de trois nourrissons âgés entre 3 et 14 mois. L'étude a débuté par une condition contrôle dans laquelle la fréquence des vocalisations des nourrissons était recueillie pendant 4 intervalles de 2 minutes. Puis, deux interventions ont été mises en place. L'intervention B (i.e., condition d'imitation vocale contingente) consistait à imiter la vocalisation que venait de produire le nourrisson. Quant à l'intervention C (i.e., condition de stimulus vocal non-contingent), il s'agissait de répéter une vocalisation produite précédemment par le nourrisson, quelle qu'ait été la vocalisation émise par le nourrisson à cet essai. Les deux interventions étaient administrées de façon aléatoire au sein d'une même session. La Figure 7 présente l'effet des interventions B et C sur la fréquence des vocalisations d'un des trois nourrissons (Holly, âgée de 3 mois). Comme on peut le voir, l'intervention B a permis d'augmenter le nombre de vocalisations du nourrisson plus que ne l'a fait l'intervention C. Ce patron de données a été observé chez les trois nourrissons d'âge différent (i.e., 3, 8 ou 14 mois). Les auteurs concluent ainsi qu'une imitation contingente des vocalisations favorise davantage la production de vocalisations chez les nourrissons qu'une émission de vocalisations non contingentes.

Les plans à traitements alternés présentent plusieurs avantages. Tout d'abord, ils ne nécessitent pas le retrait d'une intervention qui se révèle être efficace. Ensuite et grâce à la comparaison des effets de plusieurs traitements différents au sein d'une même étude, le traitement le plus

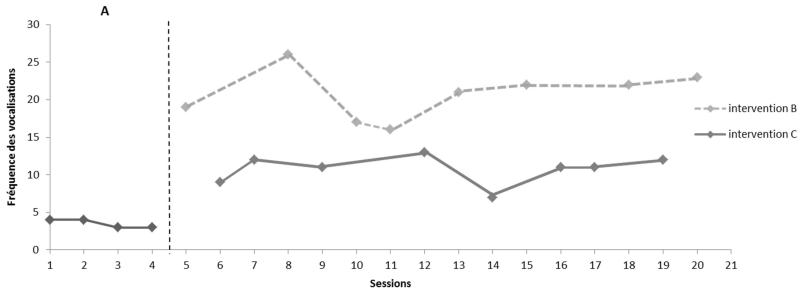


Figure 7. Plan à traitements alternés utilisé pour évaluer l'effet de deux interventions (B et C) sur la fréquence des vocalisations d'un nourrisson de 3 mois (voir Pelaez et al., 2018).

Figure 7. Alternating treatment design used to evaluate the effect of two interventions (B and C) on the vocalization frequency of a 3-month-old infant (see Pelaez et al., 2018).

efficace peut être mis en évidence plus rapidement et l'intervention la moins efficace terminée plus rapidement. Enfin, dans la mesure où on se centre sur la comparaison entre deux interventions, une ligne de base n'est pas nécessaire. Ce type de plan souffre cependant de plusieurs limites. Tout d'abord, si les différentes interventions sont trop similaires, on peut observer une généralisation de l'effet au travers des différentes interventions. Il ne sera alors pas possible de dissocier une intervention plus efficace qu'une autre. Des effets d'interférence des interventions les unes avec les autres peuvent être obtenus. L'effet de l'intervention B seule peut être différent de l'effet de l'intervention B réalisée conjointement avec d'autres interventions. Quand on utilise un plan avec plusieurs interventions, il faut prendre en compte de possibles effets d'interactions et effets d'ordre entre les traitements. Les effets observés peuvent être liés à une séquence particulière d'administration des traitements plutôt qu'à l'intervention la plus récemment administrée. Évaluer des participants supplémentaires et modifier la séquence des phases peuvent être une solution pour évaluer l'influence de ces effets d'ordre.

Les combinaisons de plans

Bien que la majorité des chercheurs utilisent les plans décrits précédemment dans leur forme classique, il est aussi possible de combiner les caractéristiques de deux ou plusieurs de ces plans dans une même étude. L'objectif est d'augmenter la force de la démonstration du lien causal entre l'intervention et le changement comportemental. En effet, montrer

les effets d'une intervention de deux ou plusieurs façons différentes renforce la confiance dans les résultats obtenus. Parmi les combinaisons classiques de plans, on peut citer l'association d'un plan avec retrait à un plan à lignes de base multiples (voir Holm et al., 2014) ou à un plan à critères changeants (voir Kowalewicz & Coffee, 2014 ; Russell, Allday, & Duhon, 2015), ou bien la combinaison d'un plan à lignes de base multiples et d'un plan à critères changeants (voir Hodges et al., 2017).

L'ÉVALUATION DES DONNÉES ISSUES DE PROTOCOLES INDIVIDUELS

L'évaluation des données a pour objectif, d'une part, de décrire les changements comportementaux observés et, d'autre part, de faire des inférences concernant cette modification. Il s'agit de s'assurer que l'effet obtenu est valide et qu'il n'est pas lié à des variations aléatoires dans la performance mesurée. Les données recueillies à l'aide de protocoles individuels fournissant un niveau de détails beaucoup plus important que les données classiquement recueillies à partir de groupes, elles nécessitent des méthodes d'évaluation différentes de celles classiquement utilisées. Dans cette perspective, deux approches sont généralement utilisées, de façon isolée ou combinée (voir Manolov & Moeyaert, 2017 ; Wolfe, Dickenson, Miller, & McGrath, 2018) : (1) l'examen visuel des données représentées graphiquement et (2) les analyses statistiques. La significativité clinique des effets obtenus est également un critère à prendre en compte lors de l'évaluation des données.

L'examen visuel des données

Historiquement, les initiateurs des protocoles individuels se sont élevés contre la nécessité de recourir aux tests statistiques pour analyser les données, en arguant que les relations causales ne sont pas démontrées par des méthodes statistiques mais grâce à l'utilisation de procédures hautement contrôlées (e.g., Skinner, 1984 ; voir aussi Horner et al., 2012). C'est de cette tradition que se sont développées les méthodes descriptives visant à évaluer visuellement la relation entre l'intervention et la variable

dépendante. La représentation graphique des données obtenues est en ce sens fondamental.

L'examen visuel des données repose sur la même stratégie que celle utilisée pour les plans inter-groupes. Il s'agit d'évaluer l'effet d'une intervention en comparant la performance obtenue dans une condition sans intervention à la performance obtenue dans une condition avec intervention. Lorsqu'on utilise un plan inter-groupes, cette comparaison est opérée en moyennant la performance d'un groupe d'individus, généralement différents, dans chacune des conditions. Or, pour les protocoles individuels, les données sont issues d'un seul individu qui a fourni plusieurs mesures dans chaque condition. Par conséquent, les méthodes utilisées pour analyser les données issues de groupes ne sont pas adaptées pour les données issues de protocoles individuels.

L'analyse visuelle des données suppose de prendre en compte simultanément 6 critères différents (voir Lane & Gast, 2014 ; Ledford, Lane & Severini, 2017 ; Parsonson & Baer, 1978 ; Kratochwill & Levin, 1992 ; Kratochwill et al., 2010) qui permettent non seulement d'examiner les données à l'intérieur d'une même phase, mais aussi entre les différentes phases.

Tableau 1. Les six critères à considérer lors d'une analyse visuelle des données issues de protocoles individuels.

Table 1. *The six criteria to consider in a visual analysis of data from single-case designs.*

Critères	Evaluation intra- et/ ou inter-phases	Définition
Niveau	intra- et inter-	Moyenne des scores dans une phase
Tendance	intra- et inter-	Pente de la ligne qui caractérise les scores dans une phase
Variabilité	intra- et inter-	Niveau de déviation des scores autour de la ligne de régression dans une phase
Recouvrement	inter-	Pourcentage de scores d'une phase qui se recouvrent avec l'étendue des scores de la phase précédente
Caractère immédiat de l'effet	inter-	Ampleur du changement entre les dernières mesures d'une phase et les premières mesures de la phase suivante
Similarité inter-phases	inter-	Degré de similarité entre des patrons de données de phases équivalentes

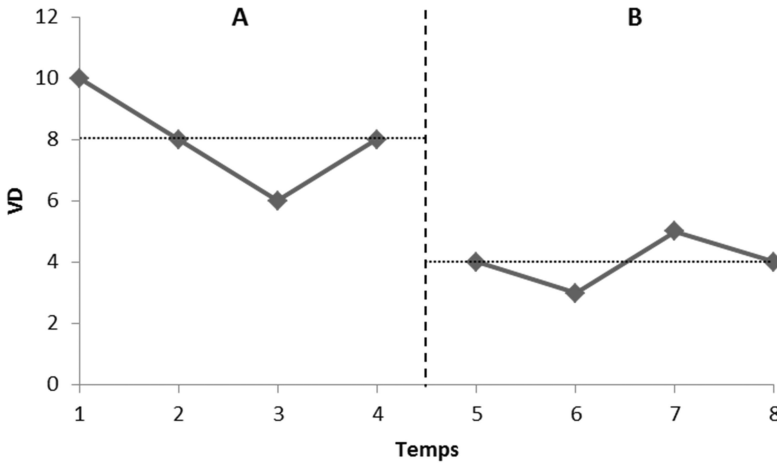


Figure 8. Exemple de performance dans un plan de type A-B où la moyenne de chaque phase est représentée par une ligne horizontale pointillée.

Figure 8. Example of performance in an A-B design where the mean in each phase is represented by a dashed horizontal line.

Tout d’abord, le critère de niveau correspond à la moyenne des données d’une phase. Après avoir calculé la moyenne des scores pour chacune des phases de l’étude, il s’agit d’évaluer si le niveau moyen change d’une phase à une autre. Dans la Figure 8, le niveau moyen de performance est de 8 dans la condition A, alors qu’il est de 4 dans la condition B, ce qui suggère que l’intervention a réduit la performance moyenne de l’individu.

Un autre critère est la tendance, correspondant à la pente de la ligne qui caractérise les données à l’intérieur d’une phase (voir Figure 9). Il s’agit d’examiner si la pente des données change d’une phase à une autre.

La variabilité correspond au niveau de variation des données autour de la ligne de régression (voir Figure 10). On peut ici prendre en compte l’étendue des scores ou calculer l’erreur-standard dans chaque phase pour la comparer avec celle de la phase suivante.

Le recouvrement (voir Figure 11) correspond au pourcentage de mesures au sein d’une phase (généralement la phase d’intervention) qui se recouvrent avec l’étendue des données au sein de la phase précédente (généralement la phase de base). Une absence de recouvrement tend à suggérer que l’intervention a provoqué un changement de comportement.

Le caractère immédiat de l’effet renvoie à l’ampleur du changement (considérée sur le niveau, la tendance ou la variabilité) entre les 3 à 5 dernières mesures d’une phase et les 3 à 5 premières mesures de la phase suivante (voir Figure 12).

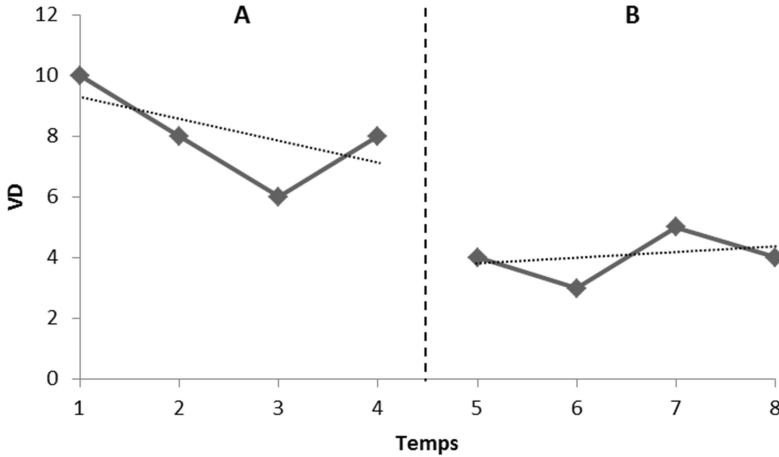


Figure 9. Exemple de performance dans un plan de type A-B où la pente de chaque phase (ici, ligne de régression linéaire) est représentée par une ligne pointillée.

Figure 9. Example of performance in an A-B design where the slope in each phase (here, the linear regression line) is represented by a dotted line.

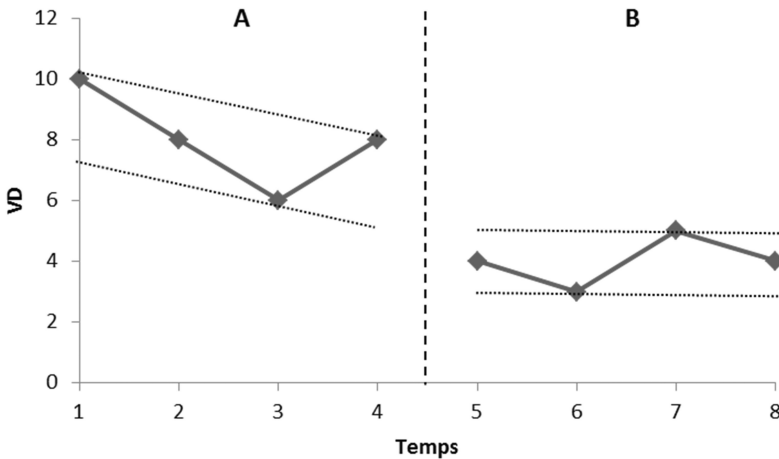


Figure 10. Exemple de performance dans un plan de type A-B où la variabilité de chaque phase (l'étendue autour de la ligne de régression) est représentée par des lignes pointillées.

Figure 10. Example of performance in an A-B design where the variability in each phase (the range around the regression line) is represented by dashed lines.

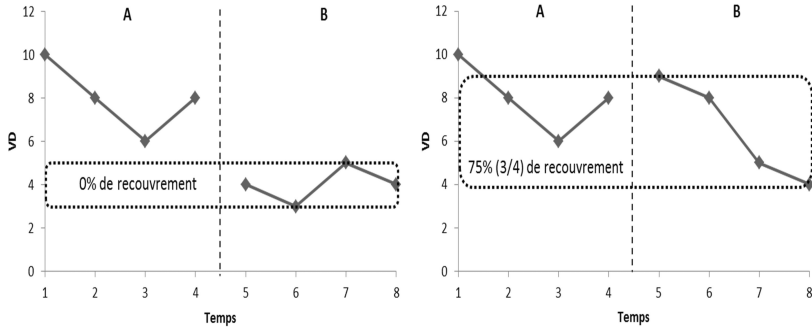


Figure 11. Exemple de performance dans un plan de type A-B où la zone de recouvrement est représentée par des lignes pointillées. À gauche, les données de la phase B ne se recouvrent pas avec celles de la phase A. À droite, 75 % des données de la phase B se recouvrent avec celles de la phase A.

Figure 11. Example of performance in an A-B design where the overlap area is represented by dashed lines. On the left, the data of phase B do not overlap with those of phase A. On the right, 75 % of data in phase B overlap with those in phase A.

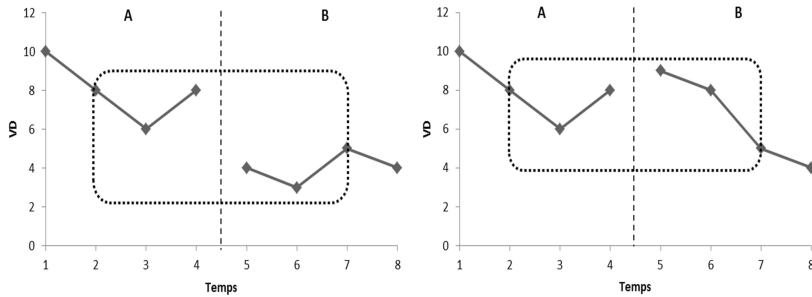


Figure 12. Exemple de performance dans un plan de type A-B où le caractère immédiat de l'effet est évalué en considérant les 3 derniers points de la phase A et les 3 premiers points de la phase B (représentés par des lignes pointillées). À gauche, le changement entre les données de la phase A et celles de la phase B est rapide. À droite, le changement entre les données de la phase A et celles de la phase B n'est pas manifeste sur les premières mesures.

Figure 12. Example of performance in an A-B design where the immediacy of effect is evaluated by considering the last 3 points of phase A and the first 3 points of phase B (represented by dashed lines). On the left, the change between the data of phase A and those of phase B is fast. On the right, the change between phase A and phase B data is not evident on the first measures.

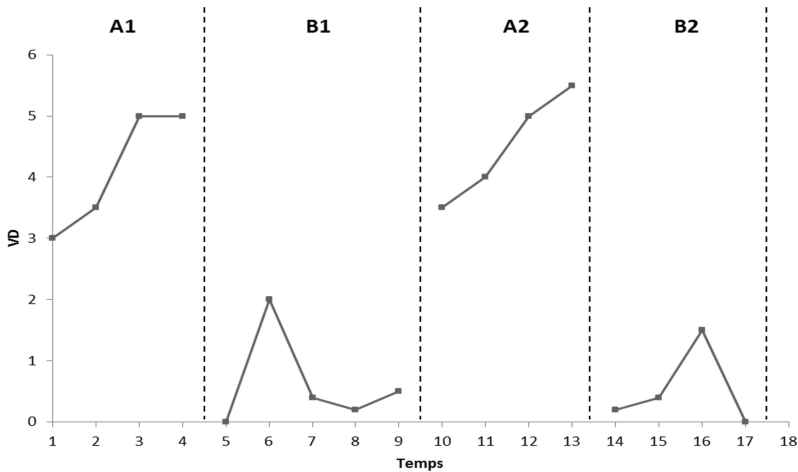


Figure 13. Exemple de performance dans un plan de type ABAB où la similarité interphases est considérée en comparant le patron de données des phases A1 et A2 à celui des phases B1 et B2.

Figure 13. Example of performance in an ABAB design where the interphase similarity is considered by comparing the data pattern of phases A1 and A2 with that of phases B1 and B2.

Enfin, le critère de similarité interphases (voir Figure 13) correspond au degré auquel les phases de conditions équivalentes sont associées à des patrons de données similaires (en termes de niveau, tendance, et variabilité). Il s'agit ici d'évaluer si les patrons de données des phases A1 et A2 se ressemblent plus que ceux des phases B1 et B2 (et vice versa).

L'analyse visuelle des données suppose non seulement de considérer ces 6 critères, mais aussi de tenir compte d'éventuelles interactions entre ces éléments (voir Kratochwill et al., 2010). Ainsi, l'évaluation des changements de niveau doit prendre en compte la variabilité des données. L'interprétation correcte d'une moyenne suppose que les données à partir de laquelle elle est calculée sont homogènes et donc peu variables. Plus les données sont variables, plus le nombre de mesures doit être important pour calculer un indice de niveau satisfaisant. De même, plus la tendance ou la pente des données est élevée, moins le niveau est pertinent pour prédire la performance de l'individu ou pour évaluer un recouvrement inter-phases. Concernant le critère de tendance, celle-ci peut être linéaire ou non. Il s'agit donc de représenter la courbe qui s'ajuste au mieux avec les données recueillies dans la phase. Une tendance (i.e., une pente) élevée nécessite de collecter de nombreux scores pour obtenir un patron de

données prédictif. Enfin, il est important de veiller à ce que la tendance ne change pas brusquement à la fin d’une phase, notamment dans le sens de l’effet attendu de l’intervention, ce qui pourrait témoigner de l’influence de variables externes sur le comportement et biaiserait les conclusions. Enfin, concernant les critères de recouvrement et du caractère immédiat de l’effet, ceux-ci ne pourront être considérés que dans le cas d’une faible variabilité des données. En effet, une variabilité élevée des mesures entraînera forcément un recouvrement quasi-total des données entre les phases. De même, des données variables à la fin d’une phase et au début de la phase suivante empêcheront la possibilité d’examiner la rapidité du changement comportemental.

Enfin, un dernier élément important à considérer lorsqu’une analyse visuelle des données issues de protocoles individuels expérimentaux est opérée, c’est la possibilité que des variables parasites fournissent une explication alternative aux résultats obtenus (voir Lane & Gast, 2014). Des indicateurs de qualité (voir Tableau 2), permettant de quantifier le niveau de rigueur méthodologique attendu pour s’assurer de la validité des données, ont ainsi été recensés afin de guider les utilisateurs dans la conduite de leur protocole (voir Horner et al., 2005 ; Kratochwill et al., 2010 ; Lane & Gast, 2014 ; Tate et al., 2016). Des échelles ont également été développées pour faciliter cette évaluation (e.g., échelle *SCED*, Tate et al., 2008 ; échelle *RoBiNT*, Tate et al., 2013).

Tableau 2. Principaux indicateurs de qualité des protocoles individuels expérimentaux (voir Horner et al., 2005 ; Kratochwill et al., 2010 ; Leford & Gast, 2018).

Table 2. Main quality indicators of experimental single-case designs (see Horner et al., 2005, Kratochwill et al., 2010, Leford & Gast, 2018).

Niveau considéré	Éléments attendus
Participants et contexte	Description précise des caractéristiques
Variable dépendante	Description opérationnelle précise Mesure quantifiable du comportement Mesurée de façon répétée dans le temps par plus d’un évaluateur Entre 3 à 5 mesures par condition Indices de fidélité et/ou accord inter-juges rapportés pour au moins 20 % des mesures (accord minimum : 80 %, Kappa = 0,60)
Variable indépendante	Description opérationnelle précise Manipulation systématique et contrôlée Indice de fidélité de la manipulation expérimentale

Niveau considéré	Eléments attendus
Validité interne	Mise en évidence de 3 démonstrations de l'effet expérimental à 3 temps différents Représentation graphique de l'effet expérimental Examen du contrôle des variables parasites
Validité externe	Réplication des effets expérimentaux (participants, contextes, et/ou matériels différents)
Validité sociale/clinique	Importance sociale et/ou clinique de la variable dépendante Importance sociale et/ou clinique de l'ampleur de l'effet de l'intervention Rapport coût-bénéfice avantageux de la mise en place de l'intervention

L'examen statistique des données

Bien que l'évaluation visuelle des données soit l'approche la plus communément adoptée pour traiter les données issues de protocoles individuels (e.g., Brossart, Parker, Olson, & Mahadevan, 2006 ; Lane & Gast, 2014 ; Ledford et al., 2017 ; Manolov & Moeyaert, 2017), une des limites majeures de cette approche est le manque de règles décisionnelles objectives pour statuer sur un changement significatif du comportement entre les différentes phases de l'étude (voir Bengali & Ottenbacher, 1998). Les utilisateurs des protocoles individuels se basent donc sur leur « sens commun » pour déterminer l'existence, de même que l'ampleur, de l'effet de l'intervention. Plusieurs chercheurs se sont alors interrogés sur la validité des inférences visuelles et ont étudié le degré d'accord inter-juges parmi des évaluateurs qui étudiaient les mêmes représentations graphiques de données issues de protocoles individuels. Ils ont ainsi montré que l'analyse visuelle des données pouvait mener à des conclusions différentes en fonction des évaluateurs sur l'effet d'une intervention. L'accord inter-juges était faible, allant de 0,50 à 0,61 (e.g., Jones, Weinrott, & Vaught, 1978 ; DeProspero & Cohen, 1979). Plus les données recueillies et le plan utilisé étaient complexes, plus le désaccord était important. Par exemple, une variabilité élevée, des changements de tendance, ou des modifications légères de niveau étaient associés aux coefficients d'accord inter-juges les plus faibles (e.g., DeProspero & Cohen, 1979). En comparant les conclusions issues d'analyses visuelles à celles provenant d'analyses statistiques, Jones et al. (1978) ont montré que ces deux méthodes fournissaient le

plus de résultats concordants dans le cas d'effets non significatifs des interventions et le plus de résultats en désaccord dans le cas d'effets significatifs des interventions. Une des raisons qui peut mener à ces désaccords est le caractère de dépendance sérielle, ou autocorrélation, des données issues de protocoles individuels (voir Bengali & Ottenbacher, 1998 ; Harrington & Velicer, 2015). En effet, les évaluations répétées auprès d'un seul individu tendent à rendre corrélées les mesures obtenues, ce qui pose problème pour bon nombre de tests statistiques classiques qui reposent sur un postulat d'indépendance des observations. Bengali et Ottenbacher (1998) ont montré que lorsque les coefficients d'autocorrélation étaient élevés, les chercheurs utilisant l'analyse visuelle concluaient davantage à une absence d'effet de l'intervention. Enfin, et pour justifier l'intérêt d'utiliser des analyses statistiques adaptées à ce type de plan, notons que les données issues de protocoles individuels sont rarement distribuées de façon normale et que des données manquantes sont souvent observées (voir Smith, 2012). La majorité des chercheurs dans ce domaine recommandent aujourd'hui l'utilisation de méthodes statistiques en complément à une analyse visuelle (e.g., Kazdin, 2011 ; Manolov, Gast, Perdices, & Evans, 2014 ; Manolov & Moeyaert, 2017 ; Wolfe et al., 2018 ; Smith, 2012).

Appliquées aux protocoles individuels, les méthodes statistiques tentent de répondre à 3 questions générales (voir Smith, 2012) : (1) L'introduction de l'intervention (i.e., la manipulation de la variable indépendante) a-t-elle provoqué un changement statistiquement significatif sur le niveau de la variable dépendante (i.e., analyse de changements de la moyenne) ? (2) L'introduction de l'intervention a-t-elle provoqué un changement statistiquement significatif sur la pente de la variable dépendante (i.e., analyse de changements de la tendance) ? Et (3) L'évolution de la variable dépendante est-elle liée à d'autres co-variables potentielles ? Par ailleurs, plusieurs auteurs (voir Harrington & Velicer, 2015 ; Heyvaert & Onghena, 2014a, 2014b ; Shadish, Hedges, Horner, & Odom, 2015) insistent sur la nécessité de rapporter les tailles d'effet afin de renseigner sur l'ampleur du changement comportemental entre la ligne de base et les phases interventionnelles. Les estimations de taille d'effets sont également fondamentales pour permettre la comparaison des effets entre études et/ou participants différents, de même que pour mettre en évidence des tendances générales au moyen de méta-analyses. Un nombre important de méthodes statistiques sophistiquées se sont développées ces dernières décennies (voir Tate et al., 2016). Nous présentons ci-dessous les principes de base de quelques-uns des principaux tests pouvant être utilisés pour traiter les données issues de protocoles individuels. Les lecteurs

intéressés pourront trouver une discussion sur les critères permettant d'orienter le choix vers une procédure statistique dans Manolov et al. (2014), ainsi qu'une liste des différents outils informatiques disponibles gratuitement pour réaliser ces analyses dans Manolov et Moeyaert (2017).

Les tests de randomisation ou tests de permutation (Edgington, 1996 ; Heyvaert & Onghena, 2014a, 2014b). Ces tests, pouvant être utilisés à la fois pour planifier le déroulement de l'étude et pour analyser les données, impliquent de répartir aléatoirement les phases d'intervention au cours de l'étude. La dépendance sérielle des données est ainsi réduite, permettant d'évaluer de façon appropriée l'ampleur de la performance (i.e., la moyenne) dans chaque condition. D'un côté, la randomisation renforce la validité interne du protocole car elle amène un contrôle statistique des variables parasites de l'étude (voir Heyvaert & Onghena, 2014a, 2014b). D'un autre côté, la randomisation est supposée améliorer la validité de la conclusion statistique dans la mesure où elle conduit à réaliser un test statistique basé sur la randomisation utilisée dans le cadre du protocole (voir Kratochwill & Levin, 2010). Le lecteur intéressé pourra trouver la procédure très détaillée et illustrée d'exemples pour conduire un test de randomisation en fonction des différents protocoles expérimentaux dans les articles de Heyvaert et Onghena (2014a, 2014b). La limite majeure de ces tests est qu'ils présentent une faible puissance statistique, de sorte que des effets fins mais systématiques peuvent ne pas être détectés (voir Haardörfer & Gagné, 2010).

Les analyses de séries temporelles (ou chronologiques) permettent de comparer les moyennes et tendances des données de chaque condition tout en prenant en compte la dépendance sérielle des mesures (voir Harrington & Velicer, 2015 ; Juhel, 2008). Elles facilitent également le calcul du d de Cohen (Cohen, 1988) pour rendre compte de la taille de l'effet de l'intervention (voir Harrington & Velicer, 2015). Ces analyses visent à déterminer si la manipulation expérimentale a produit un changement significatif entre deux phases consécutives, i.e., entre une phase A et une phase B. Toutes les phases du protocole peuvent être examinées. Par exemple, dans le cas d'un plan de type $A_1B_1A_2B_2$, des comparaisons différentes peuvent être réalisées pour chaque série de phases adjacentes (e.g., A_1 vs. B_1 , B_1 vs. A_2 , A_2 vs. B_2). Les analyses de séries temporelles permettent de détecter des effets significatifs même lorsque les effets sont fins et que la variabilité des données au sein d'une phase est élevée (voir Kazdin, 2011). Cependant, ces analyses nécessitent un nombre conséquent de mesures dans chaque phase, ce qui peut limiter leur utilisation (voir Heyvaert & Onghena, 2014b).

Les analyses multiniveaux (ou modèles linéaires hiérarchiques, modèles linéaires mixtes, e.g., Baek et al., 2014 ; Moeyaert et al., 2017 ; Rindskopf & Ferron, 2014) permettent également de quantifier des changements en termes de niveaux et de tendance (Moeyaert, Ferron, Beretvas, & Van Den Noortgate, 2014). Utiliser les analyses multiniveaux permet de répondre aux questions suivantes : (1) Quel est l'ampleur de l'effet moyen de l'intervention (quels que soient les participants et les études dans lesquelles l'intervention a été mise en place) ? (2) Est-ce que l'ampleur de l'effet de l'intervention varie pour un même participant, pour des participants différents et/ou pour des études différentes ? et (3) Des facteurs sont-ils liés à l'ampleur de l'effet de l'intervention ? Tout en prenant en compte la dépendance sérielle des mesures, les analyses multiniveaux permettent ainsi de modéliser à la fois les effets fixes (e.g., l'effet moyen de l'intervention) et les effets aléatoires (e.g., la variation de l'effet de l'intervention selon les participants ou selon les études). Des modèles à deux ou trois niveaux sont couramment utilisés pour synthétiser les données issues des protocoles individuels, notamment dans le cas des plans de type AB et des plans à lignes de base multiples (voir Moeyaert et al., 2015). Dans les modèles à deux niveaux (e.g., Ferron, Bell, Hess, Rendina-Gobioff, & Hibbard, 2009 ; Ferron, Farmer, & Owens, 2010), les données sont modélisées au travers des participants d'une même étude : les mesures (niveau 1) sont emboîtées dans chaque participant (niveau 2). Les modèles à trois niveaux (e.g., Moeyaert, Ugille, Ferron, Beretvas, & van den Noortgate, 2014 ; Owens & Ferron, 2012) vont encore plus loin dans la mesure où ils permettent de combiner les données au travers de différents participants et au travers de différentes études : les mesures (niveau 1) sont emboîtées dans chaque participant (niveau 2), eux-mêmes étant imbriqués dans une étude particulière (niveau 3). Les modèles multiniveaux semblent offrir un cadre d'analyse propice à l'analyse des données issues de protocoles individuels, qu'elles proviennent d'une seule étude, ou de différentes études qui portent sur des contenus ou problématiques similaires. Ces analyses permettent d'évaluer simultanément plusieurs séries de données individuelles et d'arriver à des conclusions précises basées sur l'ensemble des cas. Une limite potentielle de cette technique est qu'elle nécessite un nombre important d'unités de niveau 2 (30 au minimum, voir Schoebi & Reicherts, 2016, sauf si les séries de données comporte un minimum de 20 mesures), afin de s'assurer que les estimations des effets, notamment concernant la variation de l'effet de l'intervention entre les participants, soient suffisamment précises (voir Ferron et al., 2009). Le lecteur intéressé pourra trouver des détails sur la façon dont les modèles peuvent être adaptés pour modéliser

différents types d'effets (e.g., changement au niveau de la moyenne et changement au niveau de la pente) au travers de différents types de protocoles individuels expérimentaux (i.e., plans ABAB, à lignes de base multiples, à traitements alternés) dans Moeyaert et al. (2015) ainsi que dans Rindskopf et Ferron (2014).

Les analyses de recouvrement (ou de non-recouvrement). Ces analyses, directement issues de l'analyse visuelle des données, permettent de calculer facilement des tailles d'effet (voir Parker, Vannest, & Davis, 2011). Elles permettent d'apprécier l'ampleur du recouvrement des données entre des phases adjacentes. Plusieurs indicateurs ont été proposés, parmi lesquels le *PND* (*Percentage of Nonoverlapping Data*, Scuggs & Mastropieri, 1998, 2013), le *PEM* (*Percentage of data Exceeding the Median*, Ma, 2006), le *PAND* (*Percentage of All Nonoverlapping Data*, Parker, Hagan-Burke, & Vannest, 2007), le *NAP* (*Nonoverlap of All Pairs*, Parker & Vannest, 2009), ou le *PDO* (*Pairwise Data Overlap*, Parker & Vannest, 2009). Cependant, l'utilisation de ces indices est compromise lorsque les données présentent une tendance élevée, notamment au sein de la ligne de base (voir Wolery, Busick, Reichow & Barton, 2010). D'autres indices permettent de contrôler la tendance des mesures dans les phases. Il s'agit notamment de l'*IRD* (*Improvement Rate Difference*, Parker, Vannest & Brown, 2009), calculé en soustrayant le taux d'amélioration des performances observé à la ligne de base à celui obtenu lors de la phase interventionnelle, et du *Tau-U* (*Tau for Nonoverlap with Baseline Trend Control*, Parker, Vannest, Davis & Sauber, 2011) qui permet d'obtenir, au moyen de comparaisons des mesures deux à deux, le pourcentage de paires qui montrent une amélioration des performances entre la ligne de base et la phase de traitement en attribuant une pénalité lorsque des paires indiquent une dégradation des performances entre les deux conditions. Plusieurs études ont tenté d'évaluer la complémentarité des analyses visuelles et statistiques en comparant l'adéquation entre les conclusions issues d'une analyse visuelle et celles issues des méthodes de non-recouvrement (e.g., Brossart et al., 2006 ; Parker & Hagan-Burke, 2007 ; Rakap, Snyder, & Pasia, 2014 ; Wolery et al., 2010 ; Wolfe et al., 2018). Comme Wolfe et al. (2018) l'indiquent, cette approche est importante car elle contribue à améliorer l'interprétation des résultats issus des protocoles individuels expérimentaux. Dans ces études, plusieurs représentations visuelles sont générées et proposées à des participants experts (i.e., habitués à utiliser les analyses visuelles) qui doivent indiquer si les données représentées traduisent ou non un effet de l'intervention. Les évaluations des

experts sont ensuite combinées de sorte à identifier d'une part les graphiques consensuels (i.e., 80 % d'accord) qui dénotent soit un effet de l'intervention, soit aucun effet de l'intervention, et d'autre part, les graphiques non consensuels pour lesquels les conclusions des experts divergent (i.e., moins de 80 % d'accord). Enfin, les indicateurs statistiques calculés sont mis en relation avec les estimations des experts. Les études ayant comparé un ensemble différent d'indicateurs (voir Wolfe et al., 2018), il est difficile de mettre en évidence celui qui est le plus adapté pour compléter les analyses visuelles. Il semble toutefois que les indices qui prennent en compte la tendance des données (e.g., *IRD* et/ou *Tau-U*) soient ceux qui corréleront le plus avec les conclusions issues des analyses visuelles (voir Parker & Hagan-Burke, 2007 ; Wolery et al., 2010 ; Wolfe et al., 2018). Il reste désormais à déterminer dans quelle mesure ces indices sont utilisables pour tous les types de protocoles expérimentaux, et à identifier les caractéristiques spécifiques des données (e.g., tendance, variabilité, nombre de mesures) qui peuvent influencer la correspondance entre les indicateurs statistiques et les analyses visuelles (voir Wolfe et al., 2018).

Evaluation de la significativité clinique (ou validité sociale) des changements comportementaux

Outre les deux approches précédemment décrites qui correspondent à des critères dits « expérimentaux » (voir Kazdin, 2011), il est nécessaire de prendre également en compte un critère « appliqué » qui correspond à l'effet pratique que l'intervention a pu provoquer sur la qualité de vie du patient (voir Horner et al., 2005 ; Ledford & Gast, 2018 ; Spear, Strickland-Cohen, Romer, & Albin, 2013). En effet, une intervention peut avoir entraîné un changement comportemental significatif d'un point de vue expérimental sans pour autant avoir eu une répercussion clinique importante pour le patient. Plusieurs critères peuvent être pris en compte pour adresser la question de la significativité sociale et/ou clinique des données obtenues. Tout d'abord, le thérapeute peut établir avec le patient des critères de réussite. Si l'intervention atteint ce critère, le changement observé prend du sens. Ensuite, il est possible de prendre comme point de référence les seuils limites pathologiques de certains comportements et évaluer si l'intervention a permis de réduire les symptômes à un niveau en dessous de ce seuil (e.g., voir Paulus & Norton, 2016). Par exemple, dans l'inventaire abrégé de dépression de Beck (Collet & Cottraux, 1986) qui permet d'obtenir un score de dépression de 0 à 39, le seuil limite de

dépression sévère est de 16. Si une intervention permet de réduire de 10 points un score initial de 35 à cette échelle, il est possible que cette différence soit significative d'un point de vue expérimental, mais le patient sera toujours en état de dépression sévère. Enfin, il peut être aussi intéressant de considérer les coûts et bénéfices impliqués pour produire le changement comportemental. Il s'agit ici de mettre en relation les efforts fournis par le patient et par le thérapeute avec l'ampleur du changement observé (e.g., voir Haydon et al, 2010).

CONCLUSION

Les protocoles individuels expérimentaux et quasi-expérimentaux sont des outils méthodologiques puissants et efficaces qui peuvent être utilisés aussi bien en recherche fondamentale que dans la pratique clinique pour évaluer l'efficacité d'une intervention ou les progrès d'un patient. Ces plans présentent de nombreux avantages, notamment par rapport aux plans expérimentaux inter-groupes. Ainsi, ils permettent de démontrer une relation causale entre une intervention et un changement de comportement avec quelques fois plus de clarté que les plans qui utilisent des grands échantillons. Ils offrent aussi plus de flexibilité dans leur mise en place et dans l'évaluation des interventions car ils peuvent être intégrés plus facilement dans un contexte clinique. Enfin, un avantage majeur est l'obtention d'informations sur la variabilité intra-individuelle des performances de l'individu. L'individu étant évalué plusieurs fois au cours d'une même phase, la stabilité de son comportement peut être examinée, contrairement aux études de groupes qui impliquent généralement le recueil d'une seule mesure dans chaque condition pour un nombre important de participants. Bien sûr, les protocoles individuels souffrent aussi de certaines limites. Notamment, la limite la plus souvent avancée est celle du caractère généralisable des effets obtenus sur un seul individu. En effet, une intervention qui se révèle être efficace pour un individu particulier à un moment particulier peut ne pas l'être pour d'autres individus. De même, l'effet obtenu peut ne pas être répliqué si on administre à nouveau la même intervention au même individu à un moment différent. Cette limite souvent adressée à l'encontre des protocoles individuels peut cependant aussi être adressée à l'encontre des plans à groupes dont l'utilisation ne garantit pas le caractère reproductible des effets obtenus (voir Aart et al., 2015). De plus, le caractère généralisable de l'effet d'une

intervention obtenu avec un protocole individuel expérimental peut tout à fait être testé, en intégrant plusieurs individus au sein de la même étude ou en réalisant une autre étude.

En conclusion, les protocoles individuels expérimentaux et quasi-expérimentaux offrent une méthodologie propice à l'expérimentation sur le terrain. Ils fournissent un moyen efficace et scientifiquement rigoureux d'évaluer l'effet d'une intervention sur le comportement au travers de changements intra-individuels. Nous espérons que le présent article suscitera une meilleure compréhension des principes méthodologiques des protocoles individuels expérimentaux pour une utilisation plus large par la communauté scientifique de même que par les psychologues praticiens qui adhèrent à une pratique de la psychologie fondée sur la preuve.

RÉFÉRENCES

- Aarts et al. (2015). Estimating the reproducibility of psychological science. *Science*, 349.
- Baek, E.K., Moeyaert, M., Petit-Bois, M., Beretvas, S.N., Van den Noortgate, W., & Ferron J.M. (2014). The use of multilevel analysis for integrating single-case experimental design results within a study and across studies. *Neuropsychological Rehabilitation*, 24, 590-606.
- Bengali, M.K., & Ottenbacher, K.J. (1998). The effect of autocorrelation on the results of visually analyzing data from single-subject designs. *Quantitative Research Series*, 52, 650-655.
- Brossart, D.F., Parker, R.I., Olson, E.A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification*, 30, 531-563.
- Bulté, I., & Onghena, P. (2013). The single-case data analysis package: Analysing single-case experiments with R software. *Journal of Modern Applied Statistical Methods*, 12, 450-478.
- Campbell, J.M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification*, 28, 234-246.
- Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.
- Collet, L., & Cottraux, J. (1986). Inventaire abrégé de la dépression de Beck (13 items). *L'Encéphale*, 12, 77-79.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis*, 2nd ed. Upper Saddle River, N.J.: Pearson Prentice Hall.
- Cook, J.L., Rapp, J.T., & Schulze, K.A. (2015). Differential negative reinforcement of other behavior to increase wearing of a medical bracelet. *Journal of Applied Behavior Analysis*, 48, 901-906.
- Craske, M. G., & Barlow, D. H. (2014). Panic disorder and agoraphobia. In D. H. Barlow (Ed.), *Clinical Handbook of Psychological Disorders: A Step-by-Step*

- Treatment Manual (pp. 1-61). New York, NY, US: Guilford Press.
- DeProspero, A., & Cohen, S. (1979). Inconsistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis*, 12, 573-579.
- Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology* (H.A. Ruger & C.A. Bussenius, Trans.). New York: Teachers College Press (original work published 1855).
- Edgington, E.S. (1996). Randomized single-subject experimental designs. *Behavioral Research Therapy*, 84, 567-574.
- Ferron, J.M., Bell, B.A., Hess, M.R., Rendina-Gobioff, G., & Hibbard, S.T. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods*, 41, 372-384.
- Ferron, J.M., Farmer, J.L. & Owens, C.M. (2010). Estimating individual treatment effects from multiple-baseline data: A Monte Carlo study of multilevel-modeling approaches. *Behavior Research Methods*, 42, 930-943.
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6, 151-175.
- Freud, S., & Breuer, J. (1895). *Studies in Hysteria* (N. Luckhurst & R. Bowlby, Trans.). New York: Penguin (Original work published 1895).
- Ganz, J.B., & Ayres, K.M. (sous presse). Methodological standards in single-case experimental design: Raising the bar. *Research in Developmental Disabilities*.
- Gast, D. L. (2010). *Single Subject Research Methodology in Behavioral Sciences*. New York, NY: Routledge.
- Gottfredson, D. C., Cook, T. D., Gardner, F. E., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science*, 16, 893-926.
- Haardörfer, R., & Gagné, P. (2010). The use of randomization tests in single-subject research. *Focus on Autism and Other Developmental Disabilities*, 25, 47-54.
- Haydon T., Conroy M., Scott T., Sindelar P., Barber B., Orlando M. (2010). A Comparison of three types of opportunities to respond on student academic and social behaviors. *Journal of Emotional Behavior Disorders*, 18, 27-40.
- Halldorsson, B., & Salkovskis, P. M. (2017). Treatment of obsessive compulsive disorder and excessive reassurance seeking in an older adult: A Single case quasi-experimental design. *Behavioural and Cognitive Psychotherapy*, 45, 616-628.
- Harrington, M., & Velicer, W.F. (2015). Comparing Visual and Statistical Analysis in Single-Case Studies Using Published Studies. *Multivariate Behavioral Research*, 50, 162-183.
- Hendricks, L., de Kleine, R.A., Heyvaert, M., Becker, E.S., Hendricks, G.J., & van Minnen, A. (2017). Intensive prolonged exposure treatment for adolescent complex posttraumatic stress disorder: a single-trial design. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 11, 1229-1238.
- Heyvaert, M., & Onghena, P. (2014a). Analysis of single-case data: Randomisation tests for measures of effect size. *Neuropsychological Rehabilitation*, 24, 507-527.
- Heyvaert, M., & Onghena, P. (2014b). Randomization tests for single-case experiments: State of the art, state of the science, and state of the application. *Journal of Contextual Behavioral Science*, 3, 51-64.
- Hodges, A., Davis, T., Crandall, M., Phipps, L., & Weston, R. (2017). Using shaping to increase foods consumed by children with autism. *Journal of Autism Developmental Disorders*, 47, 2471-2479.

- Holm, M.B., Baird, J.M., Kim, Y.J., Rajora, K.B., D'Silva, D., Podolinsky, L., Mazefsky, C., & Minshew, N. (2014). Therapeutic horseback riding outcomes of parent-identified goals for children with autism spectrum disorder: an ABA' multiple case design examining dosing and generalization to the home and community. *Journal of Autism Developmental Disorders*, *44*, 937-947.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., Wolery, M. (2005). The use of single subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*, 165-179.
- Horner, R. H., & Kratochwill, T. R. (2012). Synthesizing single-case research to identify evidence-based practices: Some brief reflections. *Journal of Behavioral Education*, *21*, 266-272.
- Horner, R. H., Swaminathan, H., Sugai, G., & Smolkowski, K. (2012). Considerations for the systematic analysis and use of single-case research. *Education and Treatment of Children*, *35*, 269-290.
- Ingram, K., Lewis-Palmer, T., & Sugai, G. (2005). Function-Based Intervention Planning, Comparing the Effectiveness of FBA Function-Based and Non-Function-Based Intervention Plans. *Journal of Positive Behavior Interventions*, *7*, 224-236.
- Jones, R. R., Weinrott, M. R., & Vaught, R. S. (1978). Effects of serial dependency on the agreement between visual and statistical inference. *Journal of Applied Behavior Analysis*, *11*, 277-283.
- Juhel, J. (2008). Les protocoles individuels dans l'évaluation par le psychologue praticien de l'efficacité de son intervention. *Pratiques Psychologiques*, *14*, 357-373.
- Kazdin, A. E. (2011). *Single-Case Research Designs, Second Edition*. New York, NY: Oxford University Press.
- Kazdin, A.E. (1981). Drawing valid inferences from case studies. *Journal of Consulting and Clinical Psychology*, *49*, 183-192.
- Kowalewicz, E. A., & Coffee, G. (2014). Mystery Motivator: A Tier 1 classroom behavioral intervention. *School Psychology Quarterly*, *29*, 138-156.
- Krasny-Pacini, A., & Evans, J. (2018). Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide. *Annals of Physical and Rehabilitation Medicine*, *61*, 164-179.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., Shadish, W. R. (2010). Single case designs technical documentation. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf
- Kratochwill, T. R., & Levin, J. R. (Eds.). (1992). *Single-Case Research Design and Analysis: New Directions for Psychology and Education*. Hillsdale, NJ: Erlbaum.
- Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. *Psychological Methods*, *15*, 124-144.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.
- Lane, J.D., & Gast, D.L. (2014). Visual analysis in single case experimental design studies: brief review and guidelines. *Neuropsychological Rehabilitation*, *24*, 445-463.
- Ledford, J. R. & Gast, D.L. (2018). *Single Case Research Methodology, Applications in Special Education and Behavioral Sciences (3rd edition)*. New York, NY: Routledge.
- Ledford, J., Lane, J., & Severini, K. (2018). Systematic use of visual analysis for assessing outcomes in single case design studies. *Brain Impairment*, *19*, 4-17.
- Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject

- researches: percentage of data points exceeding the median. *Behavior Modification*, 30, 598-617.
- Manolov, R., Gast, D.L., Perdices, M., & Evans, J.J. (2014). Single-case experimental designs: reflections on conduct and analysis. *Neuropsychological Rehabilitation*, 24, 634-650.
- Manolov, R., & Moeyaert, M. (2017). How Can Single-Case Data Be Analyzed? Software Resources, Tutorial, and Reflections on Analysis. *Behavior Modification*, 41, 179-228.
- Melrose, S. (2015). Seasonal Affective Disorder: An Overview of Assessment and Treatment Approaches. *Depression Research and Treatment*, 1, 1-6.
- Moeyaert, M., Ferron, J., Beretvas, S. N., & Van den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of single-case experimental designs. *Journal of School Psychology*, 52, 191-211.
- Moeyaert, M., Ugille, M., Ferron, J.M., Beretvas, S.N., & Van den Noortgate, W. (2014). Three-Level Analysis of Single-Case Experimental Data: Empirical Validation. *The Journal of Experimental Education*, 82, 1-21.
- Moeyaert, M., Ugille, M., Ferron, J. M., Onghena, P., Heyvaert, M., Beretvas, S. N., & Van den Noortgate, W. (2015). Estimating intervention effects across different types of single-subject experimental designs: Empirical illustration. *School Psychology Quarterly*, 30, 50-63.
- Moras, K., Telfer, L. A., & Barlow, D. H. (1993). Efficacy and specific effects data on new treatments: A case study strategy with mixed anxiety-depression. *Journal of Consulting and Clinical Psychology*, 61, 412-420.
- Nicolas, S. (2016). *Histoire de la psychologie* (2^e éd). Paris : Dunod.
- Nock, M. K., Goldman, J. L., Wang, Y., & Albano, A. M. (2004). From Science to Practice: The Flexible Use of Evidence-Based Treatments in Clinical Settings. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43, 777-780.
- Nock, M.K., & Kurtz, S.M.S. (2005). Direct behavioral observation in school settings: Bringing science to practice. *Cognitive and Behavioral Practice*, 12, 359-370.
- Nock, M. K., Michel, B. D., & Photos, V. (2007). Single-case research designs. In D. McKay (Ed.), *Handbook of Research Methods in Abnormal and Clinical Psychology* (pp. 337-350). Thousand Oaks, CA: Sage Publications.
- Owens, C.M., & Ferron, J.M. (2012). Synthesizing single-case studies: A Monte Carlo examination of a three-level meta-analytic model. *Behavior Research Methods*, 44, 795-805.
- Parker, R. I., & Brossart, D. F. (2003). Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy*, 34, 189-211.
- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education*, 40, 194-204.
- Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: nonoverlap of all pairs. *Behavior Therapy*, 40, 357-367.
- Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single case research. *Exceptional Children*, 75, 135-150.
- Parker, R. I., Vannest, K., & Davis, J.L. (2011). Effect size in single-case research: a review of nine nonoverlap techniques. *Behavior Modification*, 35, 303-322.
- Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single case research: Tau-U. *Behavior Therapy*, 42, 284-299.
- Parsonson, B.S., & Baer, D.M. (1978). The analysis and presentation of graphic data. In T.R. Kratochwill (Ed.), *Single-Subject*

- Research: Strategies for Evaluating Change* (pp. 101-165). New York: Academic Press.
- Paulus, D.J., & Norton, P.J. (2016). Purging anxiety: A case study of transdiagnostic CBT for a complex fear of vomiting (emetophobia). *Cognitive and Behavioral Practice, 23*, 230-238.
- Pelaez, M., Borroto, A. R., & Carrow, J. (2018). Infant vocalizations and imitation as a result of adult contingent imitation. *Behavioral Development, 23*, 81-88.
- Rakap, S., Snyder, P., Pasia, C. (2014). Comparison of nonoverlap methods for identifying treatment effect in single-subject experimental research. *Behavioral Disorders, 39*, 128-145.
- Rindskopf, D., & Ferron, J. (2014). Using multilevel models to analyze single-case design data. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-Case Intervention Research: Statistical and Methodological Advances* (pp. 221-246). American Psychological Association
- Rizvi, S.L. & Nock, M.K. (2008). Single-case experimental designs for the evaluation of treatments for self-injurious and suicidal behaviors. *Suicide and Life-Threatening Behavior, 38*, 498-510.
- Russel, C.S., Allday, R.A., & Duhon, G.J. (2015). Effects of increasing distance of a one-on-one paraprofessional on student engagement. *Education and Treatment of Children, 38*, 193-210.
- Sacks, O. (1992). *L'Homme qui prenait sa femme pour un chapeau*. Paris : Seuil.
- Schoebi, D. & Reicherts M. (2016). Méthodes paramétriques: des séries temporelles avec les modèles ARIMA aux analyses multiniveaux d'analyse des cas singuliers multiples. In M. Reicherts & P. A. Genoud (Eds.), *L'Analyse du cas singulier dans la pratique et la recherche psychosociale*. (pp. 193-226). Weitraamsdorf: Edition ZKS.
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research. *Issues and applications. Behavior Modification, 22*, 221-242.
- Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education, 34*, 9-19.
- Shadish, W.R., Hedges, L.V., Horner, R.H., and Odom, S.L. (2015). *The Role of Between-Case Effect Size in Conducting, Interpreting, and Summarizing Single-Case Research* (NCER 2015-002) Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. This report is available on the Institute website at <http://ies.ed.gov/>.
- Sheehan, D. V. (1983). *The Anxiety Disease*. New York, NY: Scribner.
- Skinner, B.F. (1984). Methods and theories in the experimental analysis of behavior. *The Behavioral and Brain Sciences, 7*, 511-546.
- Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods, 17*, 510-550.
- Smith, A. H., Paulus, D. J., & Norton, P. J. (2017). Transdiagnostic assessment of anxiety symptoms using the Anxiety Disorder Diagnostic Questionnaire-Weekly Version. *Anxiety Stress Coping, 30*, 96-106.
- Spear, C.F., Strickland-Cohen, M.K., Romer, N., & Albin, R.W. (2013) An Examination of Social Validity within Single-Case Research with Students with Emotional and Behavioral Disorders. *Remedial and Special Education, 34*, 357-370.
- Stapleton, D., & Hawkins, A. (2015). Single-case research design: An alternative strategy for evidence-based practice. *Athletic Training Education Journal, 10*, 256-266.

- Steinberg, A.M., Brymer, M.J., Decker, K.B., & Pynoos, R.S. (2004). The University of California at Los Angeles Post-traumatic Stress Disorder Reaction Index. *Current Psychiatry Reports*, 6, 96-100.
- Tate, R.L., McDonald, S., Perdices, M., Togher, L., Schultz, R., & Savage, S. (2008). Rating the methodological quality of single-subject designs and n-of-1 trials: Introducing the Single-Case Experimental Design (SCED) Scale. *Neuropsychological Rehabilitation*, 18, 385-401.
- Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., . . . Wilson, B. (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 statement. *Archives of Scientific Psychology*, 4, 1-9.
- Tate, R. L., Perdices, M., Rosenkoetter, U., Wakim, D., Godbee, K., Togher, L. & McDonald, S. (2013). Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: The 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. *Neuropsychological Rehabilitation*, 23, 619-638.
- Tummers, B. (2006). DataThief III. Disponible sur le site suivant : <https://datathief.org/>
- Veale, D., Ellison, N., Boschen, M. J., Costa, A., Whelan, C., Muccio, F., & Henry, K. (2013). Development of an inventory to measure specific phobia of vomiting (emetophobia). *Cognitive Therapy and Research*, 37, 595-604.
- Wolery, M., Busick, M., Reichow, B., & Barton, E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *Journal of Special Education*, 44, 18-28.
- Wolfe, K., Dickenson, T.S., Miller, B., & McGrath, K.V. (2018). Comparing visual and statistical analysis of multiple baseline design graphs. *Behavior Modification*.
- Xin, J.F., Sheppard, M.E. & Brown, M. (2017). Brief report: Using iPads for self-monitoring of students with autism. *Journal of Autism Developmental Disorders*, 47, 1559-1567.