

## WITHIN SUBJECT DESIGNS IN THE EVALUATION OF EVIDENCE BASED PSYCHOEDUCATIONAL PRACTICES

Jackeline Joyce de Santana Santos<sup>1</sup>, Orcid: <http://orcid.org/0000-0002-4382-1066>

Ana Carolina Sella<sup>1 2</sup>, Orcid: <http://orcid.org/0000-0002-9522-8379>

Daniela Mendonça Ribeiro<sup>1</sup>, Orcid: <http://orcid.org/0000-0003-4146-5122>

**ABSTRACT.** In many countries evidence-based education (EBE) plays a central role in decision making regarding educational processes and procedures. In Brazil, EBE is still not widespread in Education, but given the gradual increase in publications and organizations related to this topic, and given its importance in the international educational scenario, it is hoped that this practice be more intensely discussed in the country, especially by professionals working in the intersection between Psychology and Education. Based on the decision making process of EBE, more specifically, on the need for a systematic evaluation of educational interventions, and having Behavior Analysis as theoretical and conceptual basis, this article has the goal of (a) presenting how the diagnostic and formative evaluations proposed by the Behavior Analysis are related to traditional ways of implementing the evaluations of psychoeducational practices and (b) demonstrating the importance of intra-subject designs as a way of evaluating and monitoring psychoeducational interventions in the quest for decisions that lead to EBE, having the published literature as its basis. The analysis made throughout this manuscript led to the conclusion that intra-subject designs can be used as evaluations in the educational practice as evidence producers.

**Keywords:** Experimental methods; evaluation; psychopedagogy.

## DELINEAMENTOS INTRASSUJEITOS NA AVALIAÇÃO DE PRÁTICAS PSICOEDUCACIONAIS BASEADAS EM EVIDÊNCIA

**RESUMO.** Em diversos países, a educação baseada em evidências (EBE) ocupa um papel central na tomada de decisões acerca dos processos e procedimentos educacionais. No Brasil, a EBE ainda não é muito difundida na educação, mas dado o aumento gradual de publicações e organizações relacionadas a esta temática e dada a sua importância no cenário educacional internacional, é esperado que esta prática passe a ser mais discutida no país, especialmente pelos profissionais que trabalham na interseção entre a psicologia e a educação. Tendo por base o processo de tomada de decisão da EBE, mais especificamente, a necessidade de avaliação sistemática das intervenções educacionais e, tendo como corpo teórico-conceitual a análise do comportamento, os objetivos deste artigo são (a) apresentar como as formas de avaliação diagnóstica e formativa propostas pela análise do comportamento se

<sup>1</sup> Universidade Federal de Alagoas, Maceió-AL, Brazil.

<sup>2</sup> E-mail: carolsella@yahoo.com.br



relacionam a formas tradicionais de implementar avaliações de práticas psicoeducacionais e (b) demonstrar a importância dos delineamentos intrassujeitos como forma de avaliação e monitoramento de intervenções psicoeducacionais na busca por decisões que levem à EBE a partir da literatura da área. As análises realizadas neste manuscrito permitem a conclusão de que delineamentos intrassujeitos podem ser utilizados como avaliações na prática educacional como produtores de evidência.

**Palavras-chave:** Métodos experimentais; avaliação; psicopedagogia.

## MODELOS INTRASUJETOS EN LA EVALUACIÓN DE PRÁCTICAS PSICOEDUCATIVAS BASADAS EN EVIDENCIA

**RESUMEN.** En diversos países la educación basada en evidencias (EBE) ocupa un papel central en la toma de decisiones acerca de los procesos y procedimientos educativos. En Brasil, la EBE aún no es muy difundida en la Educación, pero dado un aumento gradual de publicaciones y organizaciones relacionadas con esta temática y dada su importancia en el escenario educativo internacional, se espera que esta práctica pase a ser más discutida en el país, especialmente por los profesionales que trabajan en la intersección entre la Psicología y la Educación. En base al proceso de toma de decisión de la EBE, más específicamente, la necesidad de evaluación sistemática de las intervenciones educativas y, teniendo como cuerpo teórico-conceptual el Análisis de la Conducta, los objetivos de este artículo son (a) presentar como las formas de la evaluación diagnóstica y formativa propuestas por el Análisis de la Conducta se relacionan con formas tradicionales de implementar evaluaciones de prácticas psicoeducativas y (b) demostrar la importancia de los modelos intrasujetos como forma de evaluación y monitoreo de intervenciones psicoeducativas en la búsqueda de decisiones que lleven a la EBE a partir de la literatura del área. Las analices realizadas en este manuscrito permiten la conclusión de que modelos intrasujeto pueden ser utilizados como evaluaciones en la práctica educativa como productores de evidencia.

**Palabras clave:** Métodos experimentales; evaluación; psicopedagogía.

### Introduction

Evidence-based practices (EBPs) are the subject of discussion in several professional areas (e.g., Cook et al., 2014; Gambrill, 2018; Rousseau & Gunia, 2016). In Education, such practices are known as evidence-based education (EBE). The three main characteristics of EBE are: the use of the best available evidence (from research in the area of interest), the consideration of the learner's context and values, and the need for expertise by the decision maker (e.g., the teacher), which is much more determined by the years and intensity of professional training than by the years of independent experience<sup>3</sup> (Slocum et al., 2014). In addition, an important aspect of EBE is the need for a systematic and continuous evaluation of the interventions used in the place where they are adopted, i.e. there must be a partnership between science and society (e.g., organizations, families): those who decide

---

<sup>3</sup> Independent professional practice here refers to that in which professionals do not seek continuous training or supervision for their practice.

to adopt a particular EBE should produce and disseminate the adaptations and the results produced in their own context (e.g., Green, 2008; Horner et al., 2005; Slocum et al., 2014).

In Brazil, EBE is still not widespread<sup>4</sup>, but given the slow but gradual increase in publications and organizations (e.g. Alfa e Beto Institute) related to this theme and given its importance in the international educational scenario, it is expected that this practice be more discussed and used in the country, especially by professionals involved in the intersection between Psychology and Education (Kratochwill & Stoiber, 2002).

Based on the decision-making process of the EBPs, more specifically, the need for a systematic local evaluation of educational interventions, and having the Behavior Analysis as theoretical-conceptual body, which has intensely discussed its practical derivations in relation to EBPs, the objectives of this article are: (a) to present how the diagnostic and formative evaluations proposed by the Behavior Analysis are related to traditional ways of implementing the evaluations of educational process so that professionals working in the interface between Education and Psychology may understand such evaluations and use them in their practices; and (b) to demonstrate the possible importance of intra-subject designs (also called single-subject designs) as a way of evaluating and monitoring psychoeducational interventions in the search for decisions that lead to EBPs.

## Evaluation of educational interventions in behavior analysis

For Behavior Analysis, any educational intervention is, in a way, a research process because it is the duty of the person responsible for the teaching process to demonstrate that the learner has acquired skills that he did not have before (Slocum et al., 2014). The objective of the professional is to analyze the variables that interfere in the learner's behavior in order to contribute significantly to its improvement (Pressley, Graham, & Harris, 2006). It is important to emphasize, as discussed by Bogo and Laurenti (2012), that such analysis must be done in a critical way and that the good of all involved should be taken into account and not simply the correction or regulation of behaviors that are inconvenient for the *status quo*. In interventions, as in applied research, the professional analyzes behaviors that are socially important and its fostering variables and seeks experimental control<sup>5</sup> in the interventions, and yet considering the fact that these will be performed in natural environments of the behavior and, therefore, the experimental control will not be the same as that obtained in laboratory research (Baer, Wolf, & Risley, 1968).

By understanding educational interventions as processes in which objectives and priorities need to be established jointly with all stakeholders (Fawcett, 1991), designing teaching procedures, and monitoring data while implementing them so that changes can be made on the basis of the performance of the learners themselves, evaluation becomes an indispensable tool for making any decision (Christ, 2007; Kazdin, 2010). Constant and

---

<sup>4</sup> A non-systematic search on Google Scholar and Scielo on January 19, 2017, using the keywords 'evidence-based educational practices' did not result in any reference of books or articles, while using 'evidence-based education' resulted in only 23 items in Google Scholar (most of which were not published in peer-reviewed journals) and no items in Scielo. In general, from 2005 to 2010, only two articles published per year. From 2011 onwards, there was a considerable increase, reaching 12 articles in the year 2016.

<sup>5</sup> Some people may wonder about the need for experimental control in educational interventions. Our culture of utilization of trial and error in the teaching-learning process explains our lack of questioning regarding the need to know the variables that determine learning. With increasing responsibility and demand for effective results in education in many countries, verifying the relationships between intervention and results has become an important part of the teaching-learning process and this is why there is the need to demonstrate experimental control (Slocum et al., 2014; Twyman & Sota, 2009).

systematic evaluation is what provides data for the continuation or modification of the proposed interventions. Next, we will describe how evaluations are usually defined in the current context of Education and Behavior Analysis, emphasizing intra-subject designs which can provide tools for the constant, systematic and objective monitoring of educational interventions.

### **Diagnostic, formative and summative evaluations**

Usually, in the context of Education, the forms of evaluation are divided into diagnostic, formative and summative (Rabelo, 2009; Sant'anna, 2013). Diagnostic evaluation is usually carried out first (Rabelo, 2009; Sant'anna, 2013) to verify the knowledge and skills of the learner, forming the basis for individualized decision making in the learning process (Rabelo, 2009; Sant'anna, 2013).

The second form is the formative evaluation, which provides the teacher and learner with information about the learning progress in school activities<sup>6</sup> and allows the search for solutions to problems that arise during the process in an attempt to guarantee the success of the student (Rabelo, 2009; Sant'anna, 2013). This evaluation presupposes that the student is a participant in his learning process: both the students and the teachers can reflect on the data collected and suggest changes to improve the process (Rabelo, 2009; Sant'anna, 2013).

The third form of evaluation, widely used in Education, is the summative evaluation, whose main function is to classify students (Rabelo, 2009; Sant'anna, 2013). Generally, it is carried out at the end of program units, semesters and academic years, consisting of exams, tests and/or final tests, which serve to know the results of the teaching-learning process and make an overall assessment of the interim syntheses, as well as to provide elements for the teachers to make decisions about the learners' approval or disapproval (Rabelo, 2009). There is no return to the difficulties presented by the learners in the teaching-learning process; there is a conclusion whether the learners are approved or rejected according to pre-established criteria (Rabelo, 2009; Sant'anna, 2013).

Although using different terms, Behavior Analysis bases its educational practices on diagnostic and formative evaluations. If the diagnostic evaluation can be understood as a prognostic tool about the capacities, needs, interests and prerequisites of each learner at the beginning of the teaching-learning process (Rabelo, 2009; Sant'anna, 2013), this can be in many points compared to the so-called baseline, established before the beginning of any behavioral educational intervention to indicate to the educator, in a systematic, objective and consistent way, the strengths and weaknesses of each learner. If the formative evaluation is understood as "[...] an educational instrument that informs and makes a valuation of the learning process followed by the students with the objective of giving to them the opportunity, at all times, of benefit from the most appropriate educational proposals" (Zabala, 1998, p. 200), this definition resembles the evaluation made during the educational intervention process in Behavior Analysis. At each teacher-learner interaction, the effects of the educational intervention are evaluated and they guide the next interaction (Sant'anna, 2013), which allows the behavioral analyst to construct evidence of the effectiveness of his teaching procedures. As described above, summative evaluations determine, at a specific point in time, what students know about a given subject, based on a previously established standard, and such evaluations are the base for approval and disapproval. Behavioral

---

<sup>6</sup>The terms 'school activities', 'educational intervention', 'treatment', and 'independent variable' are used interchangeably here.

analysts assume that if prerequisites are established and teaching methodologies are individually programmed according to the student's rhythm, skills can be taught to any learner, provided his biological, historical, individual and cultural characteristics are taken into account (e.g., Henklain & Carmo, 2013; Skinner, 2003). Standards that are predetermined for a group and punctual evaluations depart from the premises that ground the teaching-learning process on a behavioral analytical view: the criteria that determine learning must be established for each individual and performance approaches to such criteria must be frequently and systematically assessed and also reinforced (e.g., Vargas, 2009).

From the understanding that diagnostic and formative evaluations are the ones that are closest to the behavioral perspective, the concept of baseline and some experimental intra-subject designs are presented below. We expect to demonstrate that the evaluation forms proposed by the Behavior Analysis can contribute to a meaningful teaching-learning process, leading to EBE decision making.

## The baseline

In general terms, baseline is the period in which the learner's behavior is assessed without the presence of educational intervention (Bailey & Burch, 2002; Horner & Baer, 1978). The baseline assessment may occur before the intervention (e.g., multiple baseline designs) or at various times, where intervention periods are interspersed with baseline (e.g., reversal designs), as will be described in the section on designs.

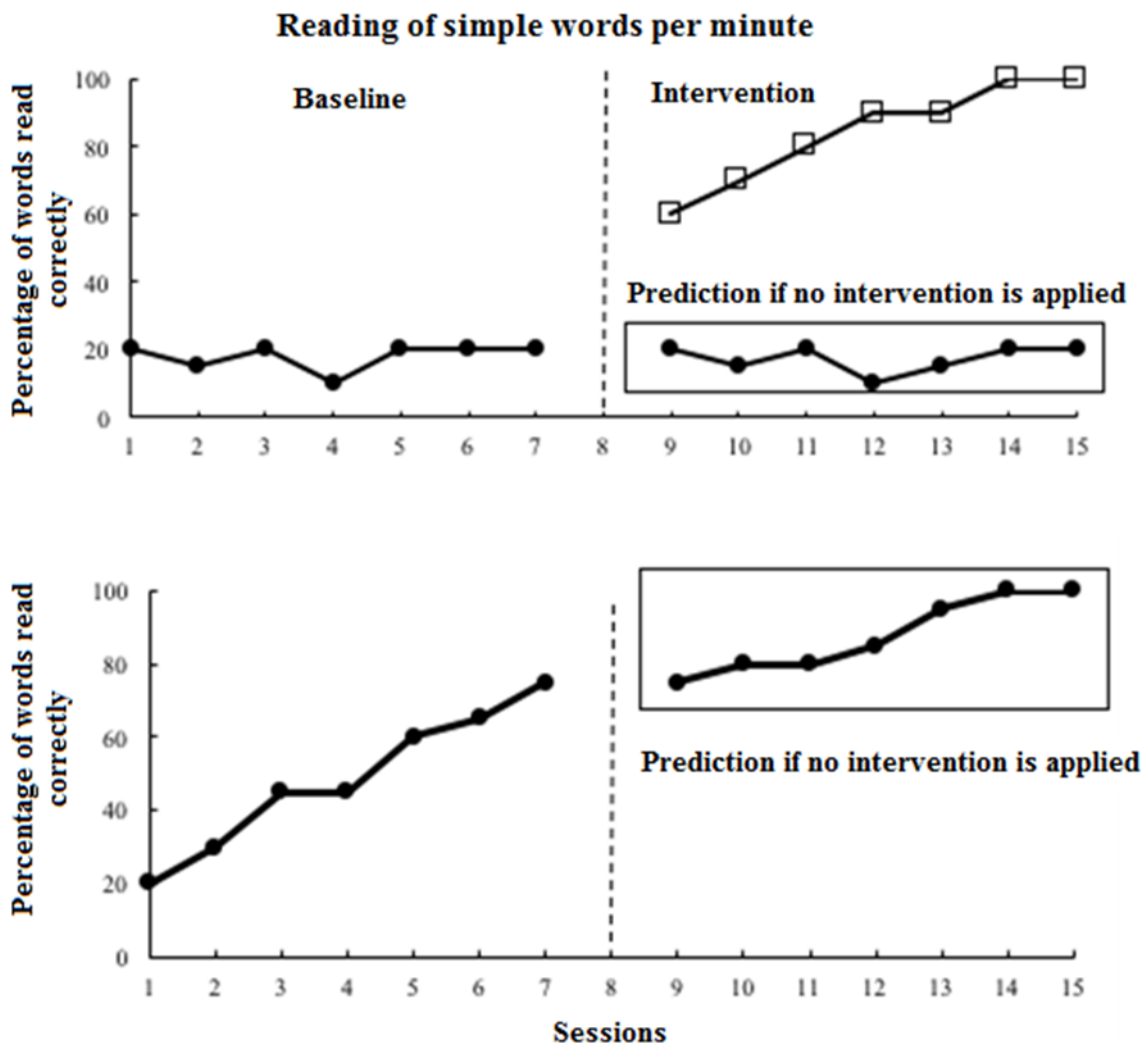
Baseline is an essential form of evaluation in Behavior Analysis, as it provides the basis for decision making regarding interventions grounded on individual learning (Kratochwill & Levin, 2015; Sampaio et al., 2008). It has the role of diagnostic evaluation, as it establishes the parameters to be compared with data resulting from the intervention. If the purpose of an intervention is to increase the frequency or duration of a given skill being demonstrated by a learner (e.g., number of words read correctly per minute), baseline data should occur less frequently than those obtained during the intervention. For behaviors that need to have their frequency or duration diminished (e.g., number of minutes talking to colleagues during the teacher's explanation), the reverse is true: during the baseline, these behaviors need to occur more frequently than in the intervention.

A well-programmed and implemented baseline helps to establish experimental control, especially in individualized educational interventions, as it provides concrete evidence of what the learner already knows and needs to learn. Without a stable and sufficient baseline, it is impossible to ensure the internal validity of interventions for learners with unique individual and sociocultural histories (Sampaio et al., 2008). The internal validity refers to the reliability of the data, that is, the possibility of saying that the intervention was responsible for learning (Baer et al., 1968; Bailey & Burch, 2002; Sampaio et al., 2008).

The baseline has three properties that make it possible to make decisions about the need for an intervention: variability, level and trend. These properties can only be evaluated from repeated measures of the dependent variable (DV - behavior of interest) without the presence of the independent variable (IV - intervention), thus ensuring that the results obtained are not simply derived from the novelty of the task, the presence of the applicator or the context in which the evaluation is being performed.

'Variability' refers to how much the behavior varies (increases or decreases) during the evaluation period. For example, if 20 evaluations of the performance of a learner on reading simple words are made, and if such evaluations show that at each evaluation he

reads correctly between 8 and 10 words per minute, it can be inferred that without a specific educational intervention, the learner will continue to read between 8 and 10 correct words per minute (8 to 10 represents the variability of this behavior). In the analysis of variability, the educator has the possibility 'to predict' what will happen to the behavior if this is not the target of a specific educational intervention (Bailey & Burch, 2002). Panel 1 of Figure 1 illustrates the predictability provided by the baseline. It is important to remember that such prediction, like any other, has its limitations, and it is more valid for the immediate future and in the context of the intervention (Kazdin, 2010).



**Figure 1.** Example of an intervention with stable baseline, prediction of behavior if no intervention is applied, and data of the hypothetical intervention (Panel 1, upper plane). Example of baseline with change towards the desired direction and its prediction without intervention (Panel 2, bottom plane).

Source: The authors.

'Trend' refers to the fact that repeated measures of behavior allow us to infer whether the behavior has a tendency to increase (ascending data), decrease (declining data) or remain the same (data with a neutral trend) concerning its frequency, duration or intensity (Kazdin, 2010; Sampaio et al., 2008). For example, if a child is required to sit for 5 minutes

to do his math activity, the time the child keeps sitting is measured several times during the baseline when math tasks are presented. If the duration of sitting increases (ascending data) during the baseline, there is no reason to implement an intervention: behavior measures indicate that this parameter has the natural tendency to change. So, the decision to implement an intervention requires that the baseline not only be stable, but also that it does not indicate a changing *trend* in the desired direction (e.g., Bailey & Burch, 2002).

The trend, along with variability, allows the educator to assume that if the learner's behavior is repeated several times without the presence of the intervention, sooner or later, he will arrive at a state of stability "[...] characterized by the absence of trend in the data (ascending or declining) and little variability in performance" (Kazdin, 2010, p. 106). If stability indicates improvement without the presence of intervention, there is no reason to implement it. The data must steadily tend towards a neutral or opposite direction to what is expected to happen in the intervention, because if they are in the same direction expected for the target behavior, it will not be possible to know if the behavior would have arrived to what was desired even without the intervention (Bailey & Burch, 2002; Sampaio et al., 2008). An example of a baseline that indicates data in the desired direction without intervention is found in the Panel 2 of Figure 1.

It is important to emphasize that "[...] stability cannot be confused with immutability" (Sampaio et al., 2008, p. 154). When dealing with behaviors, we cannot expect totally stable data because behavior is not, and will never be, completely stable. Stability refers to data with the least possible oscillation, since behavior is a changeable and ephemeral product that changes with each new interaction of the individual with the environment (Byiers, Reichle, & Symons, 2012; Sampaio et al., 2008). Stability, as presented here, is indispensable and important for the internal validity of interventions evaluated by means of intra-subject designs, that is, for the reliability of the data.

The third baseline property is the 'level' of the data. The level, along with the trend and the variability, determines whether baseline performance indicates the need for intervention. The level refers to the learner's average performance in a given experimental condition, be it such performance measured in terms of frequency, duration, latency or any other property of the behavior (Horner et al., 2005). Commonly, when it is desired that a behavior increases its values on the Y axis in the presence of the intervention, the baseline performance level should be low, and during the intervention there should be an increase in the level until reaching the criterion of performance pre-established as indicative of intervention success. For example, if it is desired to increase the frequency of math problems correctly solved by a child in a 50-minute class, it is necessary to check if the baseline level (average performance) is low enough to warrant an intervention (e.g., 2 correctly solved problems out of the 30 presented). For behaviors to have their level decreased, the opposite is true: the baseline level should be high, and there must be decreasing levels in the intervention. An example would be the case of a child who bites himself for 50 out of the 60 minutes of the observation period (baseline level), and the goal is to reach 0 (zero) minutes out of the 60 minutes of the intervention. It is important that for each target behavior, the commonly accepted levels in both the baseline and the learning criteria with the intervention be sought in the literature, while at the same time there be reflection on the children reality and on the variables that affect their behavior. This will aid in decision making about the need for intervention and in setting the performance criteria.

The following are some of the intra-subject designs, their advantages and disadvantages, and the role of the baseline and other measures of target behaviors in these

designs as ways of evaluating and producing evidence for the educational intervention process.

## Intra-subject designs

Intra-subject designs enable an individual analysis of the learners' behavior in a given reality, when an educational intervention is planned and implemented (Bailey & Burch, 2002; Christ, 2007; Pressley et al., 2006; Sampaio et al., 2008). Such designs seek to establish causal relationships between the DVs (target behaviors) and the IVs (interventions) (Bailey & Burch, 2002; Sampaio et al., 2008).

Interventions and surveys that use intra-subject designs are different from group surveys that involve the use of statistical analysis (Bailey & Burch, 2002; Kazdin, 2010; Sampaio et al., 2008). Researches using group designs analyze the target behavior and the changes in it, resulting from interventions, in relation to the group; an average is attributed to the change in performance. Learners are not seen as having specificities in the determinations of their behavior, but rather as subjects that are part of the group. The problem is that in the process of assigning learners to groups and using performance averages, the values found do not represent any of the unique learners and individual variability, which is extremely important for educational interventions, disappears (Bailey & Burch, 2002). When it comes to using research evidence for educational planning, it is important that the specificities of each learner are reflected in the data, or the educator may think that the interventions will serve any child in any situation, regardless of his characteristics.

Evaluating the behavior of the learner with an intra-subject design does not mean that it is possible to analyze the knowledge and skills of only 'one' learner at a time, but rather that the learner will be his own control, that is, it is based on his own behavior (DV) that decisions about interventions (IVs) will be made (Sampaio et al., 2008).

In order for the reader to be able to better visualize what intra-subject designs are and how to choose the best design in their conditions, depending on their context and the behavior or skill to be taught, a description of the main designs that can be used when considering individualized educational interventions is presented below. A discussion about the limits of individualized interventions will be presented after this exposition.

## Reversal design

The reversal design, known as the ABAB design, is one of the most important in the demonstration of experimental control in Applied Behavior Analysis (ABA) (Bailey & Burch, 2002). One of the first times in which this design was discussed as a way of evaluating socially important interventions was in the paper by Baer et al. (1968). Their logic consists in the possibility of baseline (A) and intervention (B) being alternated, replicated and compared with each other (Bailey & Burch, 2002; Sampaio et al., 2008). In the reversal design, after initial baseline data (phase A) show stability, trend and level to predict what would happen to the behavior if no intervention was presented, the intervention (IV) is introduced in a next phase (commonly called B), until the performance that was described as target is reached by the learner. Then, the intervention (IV) is removed and an evaluation is made to find out whether the behavior returns to the the initial baseline performance (A). This phase is considered the second baseline. If the behavior returns to the initial baseline measures, this is a first step in the confirmation that it was the intervention (IV) that generated the change in the target behavior. However, the evaluation process does not end



here; in order to have a better demonstration of experimental control and, for ethical reasons relating to avoidsuspending an educational process without the presence of the intervention that generated the expected benefits (Sampaio et al., 2008), the intervention is reintroduced (B), and it is verified whether there was a return to the desired performance. The reversal design receives this name because the baseline (A) and intervention (B) conditions are alternated (reversed) to verify the replication of the effects of the intervention on the target behaviors (Bailey & Burch, 2002; Sampaio et al., 2008). The baselines serve as diagnostic evaluations, as they provide data on the learners' performance in the absence of intervention. During the intervention, the evaluations do not cease: at each interaction of the learner with the intervention, data are recorded and, at the end of each teaching session, such data are used to decide whether the intervention should continue as it is, or it should be modified. These constant assessments that occur during the implementation of the intervention (phases B) can be classified as formative, because at each moment they determine whether the teaching-learning process is significant for the learners (Christ, 2007; Twyman & Sota, 2009).

There are several examples of classroom reversal designs focusing on three types of interventions involving groups: independent, dependent, and interdependent group contingencies (Cooper, Heron, & Heward, 2007). An example of applying interdependent contingencies that has been successful in the American educational context is the study of Barrish, Saunders and Wolf (1969). In order to teach social skills and reduce aggressive behaviors that interrupted the class, a game was developed that involved students in a competition for privileges available in almost all classrooms (more break time and first place in the snack queue). The game was held in a room of the 4<sup>th</sup> year with 24 students, and seven of those students had already been referred to the director because of problems such as leaving the seat and talking loud. For an hour, three times a week, the room was divided into two groups and the teacher marked on the board the students' disruptive behaviors during class. At the end of the class the team with lower scores would be the winner and be granted privileges. At the end of the study, there was a significant and reliable change in the students' problematic behaviors. This study was replicated in several other contexts, such as by Kellam et al. (2014) in a longitudinal study with 19 schools that showed that the use of this procedure correlated with a lower probability of drug dependence and abuse and risky sexual behaviors. It is important to emphasize that this is an example of intervention whose local evidences, from the US, have demonstrated its effectiveness. The application of this intervention in other cultural contexts should be guided by reflections about the cultural variables that determine any target behavior.

The ABAB designs can be presented in different variations, such as: (a) more than one type of intervention, making it an ABABC design, and in this variation, different interventions are represented by different letters (Bailey & Burch, 2002; Kazdin, 2010); (b) the number of conditions, resulting in an ABABABAB design, thus allowing more clarity in the verification of experimental control (Sampaio et al., 2008); and (c) inversion of conditions, resulting in a BAB (Bailey & Burch, 2002; Sampaio et al., 2008). This variation is made mainly in cases where the target behavior presents risks to the learner or to those close to him, and immediate intervention is required. In these cases, baseline conditions are short-lived, just enough to demonstrate that it was the intervention that generated the results (Bailey & Burch, 2002).

According to Bailey and Burch (2002), the reversal design is appropriate when the intervention to be performed is the presentation of an extrinsic element (e.g., receiving

compliments or tokens) or a temporary consequence (e.g., staying quiet in a corner of the classroom). This means that if the teacher considers the need to start from arbitrary reinforcers towards natural reinforcers (Henklain & Carmo, 2013; Skinner, 2003), this design may not be the most appropriate. This is because the withdrawal of the intervention should be possible and the target behaviors must be reversible, otherwise there will be not return to the state prior to the intervention (Bailey & Burch, 2002; Sampaio et al., 2008; Watson & Workman, 1981). For example, when introducing a teaching intervention (B) to acquire reading (A), it would not be possible to make the learner cease reading after the intervention is withdrawn, so it would not be feasible to use the reversal design in order evaluate reading learning (Sampaio et al., 2008). In addition to these issues, there may be ethical contraindications to this design when the behavior poses very severe risks to the learner or to those around him. For example, if intervention were a procedure to make the learner reduce the frequency of serious self-injurious behaviors such as eye-poking with the finger so that the eyeball gets damaged, reversal could result in organic damage to the learner, such as loss of vision. Therefore, it would not be ethical to use the reversal design (Bailey & Burch, 2002; Sampaio et al., 2008).

### **Multiple baseline design**

One of the first and most important descriptions of the multiple baseline (between behaviors) design was that of Baer et al. (1968). They say that this design allows the measurement and analysis of more than one behavior (DV) at a time, that is, it is possible to perform baseline measurements of more than one behavior at the same time. After baseline stability is achieved in all behaviors, the same intervention (IV) is presented for each behavior, one after the other (Kazdin, 2010; Sampaio et al., 2008). At each intervention presentation, the target behavior is evaluated. When intervention alone, with no other variable, results in a change in the target behavior (DV), that proves that the intervention generated a behavior change (Horner & Baer, 1978; Sampaio et al., 2008).

In the multiple baseline design between behaviors, the first step is to decide which will be the intervention targets. Such a choice implies considering the fact that the behaviors must be related, but independent. If they are dependent, the change in one will result in change in all, which will result in the loss of experimental control (Hartmann & Hall, 1976; Kearns, 1986). After choosing the target behaviors, data on each of these are collected repeatedly until stability is achieved on their respective baselines (Christ, 2007; Sampaio et al., 2008). Then, the intervention (IV) is introduced for one of the targets (DV), and it is necessary to wait for the achievement of performance according to the pre-established learning criterion for this behavior, along with the stability of this performance. After achieving stability at the first target and maintaining stable performance at the other baselines, the intervention is introduced for the next behavior, and so on, until all behaviors are exposed to intervention. During the process, it is necessary to collect, record and analyze data on target behaviors, whether they are at baseline or under intervention. This will allow demonstrating that it was the introduction of the IV and not any other variable that caused the change in the DVs.

In addition to the multiple baseline between behaviors, a multiple baseline can also occur between learners (three or more learners who exhibit the same behavior), between environments (the same behavior expressed by the same learner in at least three

environments), between conditions and with other variations (Bailey & Burch, 2002; Kazdin, 2010).

An example of using the multiple baseline design is the study of Harris (1986) in which behaviors related to classroom attention, productivity and realization of academic tasks were measured before and during the use of a self-monitoring intervention. The use of self-monitoring, both of attention and productivity, resulted in improvements in target behaviors.

Multiple baseline designs are appropriate in cases where it is not possible, ethical or practical to make a reversal to demonstrate experimental control, in the case of behaviors such as speech or writing or severe behavioral problems (Bailey & Burch, 2002; Kazdin, 2010). They are also appropriate when the goal is to check the effects of a single intervention under different conditions. Moreover, they allow the intervention to be applied gradually to one situation at a time by the persons responsible for the intervention (Kazdin, 2010). Multiple baseline designs also allow the introduction of increasingly natural reinforcers, with their respective monitoring. Disadvantages include the fact that, since it is necessary to wait for the baselines to be stable, the introduction of the IV may take an extended period of time when the behaviors of interest (DVs) do not receive intervention (Bailey & Burch, 2002; Kazdin, 2010). In addition, if the DVs are interdependent, the implementation of the intervention over a DV will lead to changes in all DVs, dissolving the experimental control, since the *sine qua non* condition for demonstration of experimental control with this design is that the DVs change only when the intervention (IV) is implemented (Bailey & Burch, 2002; Kazdin, 1982).

### Changing criterion design

Changing criterion designs are used when the learners already have a given skill, but it is necessary to increase their frequency, duration, or intensity. They can also be used when it is necessary that a given behavior decrease systematically. For example, if a learner correctly reads only 20 words per minute and he is expected to read 70, it is unreasonable to think that reading will go from 20 to 70 words per minute without intermediate steps. The changing criterion design allows such intermediate steps to be predicted, implemented and monitored. On the other hand, if it is necessary to reduce the frequency or duration of a given behavior, for example, if the learner asks to go to the bathroom every 5 minutes and it would be more appropriate for this to occur only once an hour, this should happen gradually to increase the likelihood that the educational intervention succeed.

The changing criterion design starts with a baseline that provides the data on the occurrence of the target behavior prior to intervention. The person responsible for the intervention establishes the ultimate goal and the sub-phases are delineated, with gradual performance criteria that will allow the final behavior to be achieved through successful steps (Bailey & Burch, 2002; Sampaio et al., 2008).

According to Sampaio et al. (2008), although similar to those previously described, the changing criterion design has specificities that differ from the others: different from the reversal design, the intervention need not be withdrawn or temporarily suspended to demonstrate its effect on the DV and, different from the multiple baseline design, the intervention is implemented with a single behavior.

Given the requirement for the implementation of a new sub-phase after a performance criterion is reached and at the same time the need for experimental control, it is necessary that some parameters always be present in the changing criterion design: (a) at least two

sub-phase changes throughout the design must be made; (b) before formulating a new performance criterion for the next sub-phase, the stability of the current sub-phase is of utmost importance (only when stability is achieved in each sub-phase it is that the sub-phase serves as the baseline for the next sub-phase); and (c) the defined criteria cannot go far beyond what the learner can already do independently, because he can take a long time to reach them. Such criteria cannot be undemanding, becoming something superfluous and without social relevance for the learner. Steps that require new learning must be found, without being so costly to the point of generating frustration for learners (Bailey & Burch, 2002; Sampaio et al., 2008). This design allows the learner's pace to determine the next steps, what is much more likely to generate performances that are independent of extrinsic reinforcers. An example of a changing criterion design used in a school environment was the one presented in the study by Mruzek, Cohen and Smith (2007) whose objective was to reduce the frequency of challenging behaviors, including tantrums, antisocial vocalizations and physical aggression, and also increasing the number of hours in which the participants followed the proposed rules of two 9- and 10-year-old students. Behavioral contracts containing rules to be followed were developed in collaboration with the participants. Behavioral contracts with criteria that increased rule-following hours were effective for both participants.

There are other types of intra-subject designs, however, these are the most used in school practice and in educational interventions. For those interested in learning more about intra-subjects designs, especially in education, we recommend books like Fraenkel, Wallen and Hyun (1993) and Kratochwill and Levin (2015).

## Final considerations

EBPs have played an important role in several countries concerned with establishing effective teaching practices. In Brazil, they have been increasingly publicized, especially in health areas (Brasil, 2013), but it is hoped that Education starts to consider such practices, although in a gradual manner. An example of such consideration was the launching of the Evidence Platform by the Ministry of Education and the Secretariat of Basic Education that proposes that educational technologies be evaluated based on their evidences.

One of the most important characteristics of the EBPs is the constant evaluation of the effects of educational interventions on the learning of students in their local context. Without social relevance, the use of scientific knowledge is not justified: "[...] if we want more evidence-based practices, we need more evidence based on practice" (Green, 2008, p. i23). It is necessary that the decision to adopt practices evaluated by science results in generalized improvements in the local context. Behavior Analysis and its intra-subject designs may bring important contributions to this process. The diagnostic assessment provided by the baseline of such designs, not only ensures that essential skills be present in teaching planning, but also increases learner's motivation because he will not be taught skills already existing in his repertoire. besides the motivational factor, baseline data provide the teacher with a economy of time spent with teaching because the teacher will focus his planning on skills that need to be taught rather than following a rigid schedule imposed by standards that do not take into account the individual (Pereira, Marinotti, & Luna, 2004). Data of formative evaluations obtained during the interventions provide the educators with

evidence, at each new interaction with the learners, to make decisions about the continuity or adaptation of their interventions. This type of assessment makes it possible that, after two or three interactions without change in the target behavior, educators act in the search for the students' learning (Bailey & Burch, 2002; Rabelo, 2013; Sampaio et al., 2008).

Although the intra-subject designs have their importance, it is necessary to discuss the feasibility of using these designs in a context where there is overcrowding in classrooms, lack of physical structure and of financial resources for teachers to carry out their activities and little investment in continuing education of teachers, among many other problems (e.g., Brock & Schwartzman, 2005; David, Silva, Ribeiro, & Lemes, 2015; Gatti, 2010). However, in the contexts in which these designs are adopted, there is a partnership of behavior analysts with local society, in which the so-called participatory research (Green, 2008) comes into force, producing "[...] concrete knowledge contextualized into the historical and social reality" (Bogo & Laurenti, 2012, p. 966). Without these partnerships, it will be difficult for teachers to implement such practices in their daily lives, especially without setting relevant goals.

It is also important to note that, like any other research methodology that may be used in the practice, intra-subject designs are not applicable to all educational problems. As Henklain and Carmo (2013) argue, there are many ways and approaches to educational issues. The behavioral approach has important contributions to offer to educators; however, the choice of designs and interventions will depend on the nature of their educational problem, that is, the variables involved in that specific context. In short, if "[...] pedagogical practices and teaching procedures are only valuable to the extent that they generate better learning" (Pereira et al., 2004, p. 12), reflections should be made in order to find ways to generate such learning.

## References

- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1(1), 91-97.
- Bailey, J. S., & Burch, M. R. (2002). *Research methods in applied behavior analysis*. Thousand Oaks, CA: Sage.
- Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, 2(2), 199-124.
- Bogo, A. C., & Laurenti, C. (2012). Análise do comportamento e sociedade: implicações para uma ciência dos valores. *Psicologia: Ciência e Profissão*, 32(4), 956-971.
- Brasil. (2013). *Saúde baseada em evidências*. Recuperado de: <http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/periodicos>
- Brock, C., & Schwartzman, S. (2005). *Os desafios da educação no Brasil*. Rio de Janeiro, RJ: Nova Fronteira.

- Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology*, 21(4), 397-414.
- Christ, T. J. (2007). Experimental control and threats to internal validity of concurrent and nonconcurrent multiple baseline designs. *Psychology in the Schools*, 44(5), 451-459.
- Cook, B., Buysse, V., Klingner, J., Landrum, T., McWilliam, R., Tankersley, M., & Test, D. (2014). Council for exceptional children: standards for evidence-based practices in special education. *Teaching Exceptional Children*, 46(6), 206.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, N.J.: Pearson/Merrill-Prentice Hall.
- David, C. M., Silva, H. M. G., Ribeiro, R., & Lemes, S. S. (Orgs.). (2015). *Desafios contemporâneos da educação*. São Paulo, SP: Cultura Acadêmica.
- Fawcett, S. B. (1991). Social validity: a note on methodology. *Journal of Applied Behavior Analysis*, 24(2), 235-239.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (1993). *How to design and evaluate research in education*. Nova York, NY: McGraw-Hill.
- Gambrill, E. (2018). Contributions of the process of evidence-based practice to implementation: educational opportunities. *Journal of Social Work Education*, 1-13. Recuperado de: <https://doi.org/10.1080/10437797.2018.1438941>
- Gatti, B. A. (2010). Formação de professores no Brasil: características e problemas. *Educação e Sociedade*, 31(113), 1355-1379.
- Green, L. W. (2008). Making research relevant: If it is an evidence-based practice, where's the practice-based evidence? *Family Practice*, 25(suppl. 1), i20-i24.
- Harris, K. R. (1986). Self-monitoring of attentional behavior versus self-monitoring of productivity: effects on on-task behavior and academic response rate among learning disabled children. *Journal of Applied Behavior Analysis*, 19(4), 417-423.
- Hartmann, D. P., & Hall, R. V. (1976). The changing-criterion design. *Journal of Applied Behavior Analysis*, 9(4), 527-532.
- Henklain, M. H. O., & Carmo, J. S. (2013). Contribuições da análise do comportamento à educação: um convite ao diálogo. *Cadernos de Pesquisa*, 43(149), 704-723.
- Horner, R. D., & Baer, D. M. (1978). Multiple-probe technique: a variation of the multiple baseline. *Journal of Applied Behavior Analysis*, 11(1), 189-196.
- 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(2), 165-179.

- Kazdin, A. E. (1982). *Single-case research designs*. New York, NY: Oxford University Press.
- Kazdin, A. E. (2010). *Single-case experimental designs: methods for clinical and applied settings*. New York, NY: Oxford University Press.
- Kearns, K. P. (1986). Flexibility of single-subject experimental designs. Part II: design selection and arrangement of experimental phases. *Journal of Speech and Hearing Disorders*, 51(3), 204-214.
- Kellam, S. G., Wang, W., Mackenzie, A. C. L., Brown, C. H., Ompad, D. C., Or, F., ... Windham, A. (2014). The impact of the good behavior game, a universal classroom-based preventive intervention in first and second grades, on high-risk sexual behaviors and drug abuse and dependence disorders into young adulthood. *Prevention Science: The Official Journal of the Society for Prevention Research*, 15(01), 6-18.
- Kratochwill, T. R., & Levin, J. R. (Orgs.). (2015). *Single-case research design and analysis (psychology revivals): New directions for psychology and education*. New York, NY: Routledge.
- Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: Conceptual foundations of the procedural and coding manual of division 16 and the society for the study of school psychology task force. *School Psychology Quarterly*, 17(4), 341-389.
- Mruzek, D. W., Cohen, C., & Smith, T. (2007). Contingency contracting with students with autism spectrum disorders in a public-school setting. *Journal of Developmental and Physical Disabilities*, 19(2), 103-114.
- Pereira, M. E. M., Marinotti, M., & Luna, S. V. (2004). O compromisso do professor com a aprendizagem do aluno: contribuições da análise do comportamento. In M. M. C. Hübner & M. Marinotti (Orgs), *Análise do comportamento para a educação: contribuições recentes* (p. 11-32). Santo André, SP: ESETec - Editores Associados.
- Pressley, M., Graham, S., & Harris, K. (2006). The state of educational intervention research as viewed through the lens of literacy intervention. *British Journal of Educational Psychology*, 76(1), 1-19.
- Rabelo, E. H. (2009). *Avaliação: novos tempos, novas práticas* (8a ed.). Rio de Janeiro, RJ: Vozes.
- Rabelo, M. L. (2013). *Avaliação educacional: fundamentos, metodologia e aplicações no contexto brasileiro*. Rio de Janeiro, RJ: SBM.
- Rousseau, D. M., & Gunia, B. C. (2016). Evidence-based practice: the psychology of EBP implementation. *Annual Review of Psychology*, 67(1), 667-692.
- Sampaio, A. A. S., Azevedo, F. H. B., Cardoso, L. R. D., Lima, C., Pereira, M. B. R., & Andery, M. A. P. A. (2008). Uma introdução aos delineamentos experimentais de sujeito único. *Interação em Psicologia*, 12(1), 151-164.

- Sant'anna, I. M. (2013). *Por que avaliar? Como avaliar? Critérios e instrumentos*. Rio de Janeiro, RJ: Vozes.
- Skinner, B. F. (2003). *The technology of teaching*. Cambridge, MA: B. F. Skinner Foundation.
- Slocum, T. A., Detrich, R., Wilczynski, S. M., Spencer, T. D., Lewis, T., & Wolfe, K. (2014). The evidence-based practice of applied behavior analysis. *The Behavior Analyst*, 37(1), 41-56.
- Twyman, J. S., & Sota, M. (2009). Identifying research-based practices for response to intervention: Scientifically based instruction. *Journal of Evidence-Based Practices for Schools*, 9, 86-101.
- Vargas, J. S. (2009). *Behavior analysis for effective teaching*. New York, NY: Routledge.
- Watson, P. J., & Workman, E. A. (1981). The non-concurrent multiple baseline across-individuals design: An extension of the traditional multiple baseline design. *Journal of Behavior Therapy and Experimental Psychiatry*, 12(3), 257-259.
- Zabala, A. (1998). *A prática educativa: como ensinar*. Porto Alegre: Artes Médicas Sul. Recuperado de: <http://docslide.com.br/documents/zabala-a-pratica-educativa-como-ensinar-1998.html>

Received: Aug. 12, 2017  
Approved: Sep. 27, 2018

---

*Jackeline Joyce de Santana Santos*: is a student of the Pedagogy course at the Universidade Federal de Alagoas. She has experience in Education, with emphasis on Special Education, mainly researching the following themes: Behavioral Analysis, Autism Spectrum Disorder and Evidence Based Practices. She currently works as a trainee for the Secretaria Municipal d Educação do Estado de Alagoas (SEMED-AL).

*Ana Carolina Sella*: is certified by the International Certification Board of Behavioral Analysts at the PhD level (BCBA-D). She received her PhD from the Universidade Federal de São Carlos, after a sandwich training at the University of Kansas, where she returned to complete a second master degree in Applied Behavioral Sciences. She then completed her post-docin Behavioral Psychology for Autism Spectrum Disorders at Nebraska Medical Center. She is currently a professor at the Universidade Federal de Alagoas. Her areas of expertise include teaching of basic repertoire for children with developmental delay, language teaching, instructional design, curricular adaptations, and original character of behavior.



*Daniela M. Ribeiro:* is a post-doctorin Psychology; PhD and master in Special Education and graduated in Pedagogy from the Universidade Federal de São Carlos (UFSCar). She is a lecturer at the Education Center and the Graduate Program in Education at the Universidade Federal de Alagoas (UFAL) and researcher at the Instituto Nacional de Ciência e Tecnologia sobre Comportamento, Cognição e Ensino (INCT-ECCE). The focus of her research is on the teaching of verbal and academic repertoires for children with Autistic Spectrum Disorder (ASD).