Best Practices in Utilizing the Changing Criterion Design

Liesa A. Klein · Daniel Houlihan · James L. Vincent ·
Carlos J. Panahon

Published online: 8 January 2015
© Association for Behavior Analysis International 2015

Abstract The changing criterion design (CCD) has been a recognized format of single-case research for four decades. Published examples of the CCD have been limited and the structure of the design used in the literature has varied to a degree that might engender confusion. This review examines the structure of CCD studies published to date to identify prior implementation practices and identify best practices for future use.

Keywords Single-case · Changing criterion design · Best practices

The changing criterion design (CCD) utilizes step-wise benchmarks for manipulating a dimension (i.e., accuracy, frequency, duration, latency, or magnitude) of a single behavior already present in an individual's repertoire (Cooper et al. 2007). The design has been used with behaviors where an immediate, considerable increase or decrease may be difficult to achieve or undesired; therefore, gradual shifts toward a desired goal are applied.

The CCD was first described in two books published in 1971: one a short description in a book by Hall, and the other a brief study reported by Weis and Hall which used the design to reduce daily cigarette smoking by a college student. A few subsequent studies utilized this design (i.e., Brown et al. 1974; Wallace and Davis 1974) prior to the seminal article by Hartmann and Hall (1976), which described the CCD in more detail and noted its similarity to a procedural description by Sidman (1960, pp. 254–256) in reference to the evaluation of irreversible behavior. Hartmann and Hall saw the CCD as a

type of *multiple baseline design* that could be used with a shaping procedure. In response to the CCD as a *multiple baseline design*, Cooper et al. (2007) point out that the behavior being evaluated in a CCD is intentionally interdependent (unlike a multiple baseline design) which can lead to an exaggerated perception of experimental control. Conceptualizing this as an extension of the multiple baseline design adds confusion. Consequently, the CCD is best suited for the acceleration and deceleration of a behavior, but is versatile in the behaviors and populations it can be applied to. Despite this, the CCD has only rarely been mentioned as a research or intervention tool since Hartmann and Hall's article.

Though not as experimentally robust as other designs (e.g., the multiple baseline design, which demonstrates experimental control through staggered, independent baselines), the CCD does achieve acceptable control through baseline logic and appropriate manipulation of design structure (see Table 1 for a simplified presentation of baseline logic of the CCD). With the CCD, a baseline serves as a measurement between the absence and presence of the independent variable, as do each criterion and its subsequent criterion. Therefore, when stable responding occurs at each step, (1) predictions about future responding can be made with each criterion, and (2) replications of intervention effect are observed repeatedly. Verification can be accomplished by varying the phase length or by reverting briefly to a previous criterion or baseline level (Hall and Fox 1977). Although not a requirement of this design, reverting to a previous criterion, or mini-reversal, can considerably increase experimental control by demonstrating bidirectionality of the behavior under manipulation (e.g., Barlow and Hersen 1984; Kazdin 1982, 2011). Implementation of the mini-reversal strategy in a clinical intervention to decrease separation anxiety in a child by Flood and Wilder (2004) showed a nearly perfect matching of clinical utility and experimental rigor within a rather subtle intervention utilizing the CCD. As with all single-case designs, data

L. A. Klein · D. Houlihan (✉) · J. L. Vincent · C. J. Panahon
Minnesota State University, Mankato, Mankato, MN, USA
e-mail: daniel.houlihan@mnsu.edu

Table 1 Achievement of experimental control within the CCD

Baseline logic	
Prediction 1:	The behavior will change to meet a criterion when implemented. Each criterion implemented (e.g., reversal or continued progression) serves as a hypothesis the behavior will follow.
Verification:	(1a) Vary phase magnitude and, or (1b) Implement a mini- (return to previous criterion) or full (return to baseline) reversal to demonstrate the behavior conforms to meet criterions, despite variations.
Prediction 2:	Behavioral responding will remain steady if the criterion remains stable.
Verification:	(2) Vary phase length to demonstrate that stable behavioral responding occurs as long as the criterion implemented is unchanged.
Replication	(3) Three or more criterion changes allows for repeated observation of intervention effects.
Other requirements	(4) Allow noncompulsory responding (e.g., avoid forced follow through). (5) Unrestricted responding (e.g., allowance to reach or surpass planned criterion(s) without coercion). (6) Consistent materials and behavior across criterions (e.g., avoid changes in materials that may require differing skills (e.g., subtraction to multiplication).

should guide alterations; however, variation in phase length and magnitude with more criterions rather than fewer should be attempted whenever possible to increase internal validity (Hartmann and Hall 1976). This sometimes puts clinical utility (which targets gradual change) and experimental rigor (which emphasizes robust and discrete changes in behavior) at odds. Phase length should also vary across criterions in order to (1) allow stable responding to occur (Hartmann and Hall) and (2) demonstrate that desired responding occurs as long as each criterion is in place (Cooper et al. 2007). Phase criterions should be large enough in magnitude for detectable changes and still be achievable for the individual's repertoire (Hartmann and Hall). Within these confines, variations of phase magnitude should be employed to demonstrate that changes in responding correspond to the differing criterions applied. Hartmann and Hall suggest that smaller criterion changes are necessary for behaviors resistant to change. Alberto and Troutman (2006) have suggested that the first criterion implemented can be chosen in one of four ways: using the mean of baseline, halving the mean of baseline, using the highest or lowest point at baseline to establish a repertoire range, or using a professional estimate from someone familiar with both the subject and the behavior targeted for change. The basic approach of the first three suggestions dilutes the importance of adjusting magnitude based on individual ability, and therefore the fourth suggestion of

professional estimate should be preferred in most situations. The more criterion shifts applied, the more responding is demonstrated to be controlled by these criterions. A minimum of two criterion shifts has been previously suggested (Kazdin 1982, 2011), while a more recent paper has called for a minimum standard of three criterion shifts (Kratochwill et al. 2013) which would provide opportunity for variation of phase magnitude. No further guidance on what is adequate for these variables has been provided (e.g., acceptable phase magnitude, length, mini-reversals, or number of criterion shifts needed) as each case will vary according to individual capability and resource restriction (e.g., time). Lastly, the opportunity to meet the final goal must be presented across criterions in order to demonstrate functional control (Roane et al. 2011). If this opportunity is not presented, then responding is being tailored to criterions rather than demonstrating the criterion as controlling behavior; this is a critical distinction.

Two recent variations of the changing criterion design have been suggested by McDougal. The *range-bound changing criterion design* variation consists of an upper and lower criterion for each sub-phase (McDougall 2005). The rationale for two criterions is the establishment of a maximum and minimum level of performance expected (1) to prevent injury or prevent gains toward the objective too quickly, which may reduce likelihood of long-term change; and (2) to increase experimental control through adherence to two criterions as

opposed to one. Advantages of this design for an exercise program are clear in their ability to deter excessive activity that may lead to injury. Likewise, when reducing caffeine consumption or when constricting calorie intake, a range might lead to more accurate reporting of intake, as well as more comfortable transitions between reset criterions. However, this design reduces the design integrity through limiting responding. The magnitude of each phase should be based on responding in the previous sub-phase and professional judgment. The basis for this determination still lacks clarity. Thus, when restricted responding is in place it is not representative of the individual's ability and does not lead the experimenter or therapist to an appropriate criterion shift. If a criterion is shifted five data points up to increase a behavior, experimental control is demonstrated when responding increases. If the individual meets their desired outcome almost immediately after intervention implementation and is able to maintain that goal, the inappropriate intervention was used and a poor assessment of the individual's capability and or motivation was conducted.

The second variation, *the distributed criterion design* (DCD), is proposed to manipulate one behavior across interdependent but discrete contexts or interdependent behaviors in one context (McDougall 2006). Gradual criterion shifts are not required in this variation as it is not a design targeting increasing or decreasing a dimension of behavior, but rather it functions as a behavior or response management system (e.g., self-management or self-monitoring). This management system has been demonstrated by McDougall using an overarching fixed dependent variable; however, the design could be implemented with a non-fixed variable as well. Although the sequential changing of multiple behaviors is often desirable (especially in echobehavioral interventions), and it is correct that manipulating rates of one behavior will necessarily impact other functionally similar behaviors (Delprato 2002), the value of this design is questionable. The multiple baseline design is a poor model for the DCD in that it is significantly weakened by interdependence of baselines or results, whereas interdependence is an established expectation of the DCD. It should also be noted that the example presented by McDougall is essentially the same design as the *A-B with replication design* described by Harris and Jenson (1985) and is similarly weakened by the necessity of predetermined criterions, reactive trends, and predictions in behavioral performance that must predate the actual behavior occurring. It is a fundamentally weak design that adds little to the literature in its present form.

The features of the CCD suggest its function as a research tool or an intervention tool; however, due to the systematic nature of the CCD, these descriptors become mutually exclusive. If contingencies were removed from step-wise criterions, one could still not rule out sensory consequences as a result of goal reaching in most instances (e.g., automatic

reinforcement). In other words, an individual may experience positive feelings in response to meeting a criterion, thus reinforcing the behavior independently from any prescribed reinforcer or reward. Further, if automatic reinforcement were not present in the beginning criterions, it would arguably become a classically conditioned response when paired with contingent reinforcement or punishment. Therefore, this extraneous variable precludes the CCD from being predominantly a research tool and qualifies it as an intervention tool or research/intervention tool or in some circumstances (e.g., self-monitoring) an intervention in and of itself.

The CCD is a flexible, basic design versatile in terms of applied population, target behavior, and setting, while also capable of demonstrating adequate experimental control. However, since its primary presentation in 1976, the CCD has been infrequently used as a research/intervention tool. The purpose of this paper is to examine the CCD research that has been conducted and characteristics of its use and implementation to date.

Method

Studies containing a CCD were identified through a search of *PsycINFO*, *ERIC*, and *PubMed* databases using the key word *changing criterion* as well as articles referencing Hartmann and Hall (1976). The introduction and discussion section of each article was examined to locate other studies utilizing a CCD. Articles containing both original research data and a statement that a CCD was used were included in this study. All articles identified that were published prior to January 2014 were included. A pilot run of this study was conducted with ten articles to better shape data collection and variables to be included.

Changing Criterion Design Characteristics

Data collected was specific to participants with whom a CCD was applied and only data directly connected with its implementation. For example, when a study collected baseline, implemented a treatment, and then implemented a changing criterion separate of the treatment, only the information related to the changing criterion portion was coded.

Population, Setting, and Therapist

Participants Data on participant age and diagnoses were collected. Participant age was classified as child (1 to 18 years) or adult (19 and above). Diagnoses were classified as medical (e.g., diabetes, hyperfunctional dysphonia) or physical (e.g., visual or hearing impairment), learning disability (LD), Attention Deficit Hyperactivity Disorder (ADHD), pervasive developmental disorder (e.g., Autism Spectrum Disorder),

Mental Retardation, emotional disorder (e.g., emotional behavioral disorder, separation anxiety disorder), other (e.g., schizophrenia, bulimia), and none or not reported.

Settings The settings where the CCD was applied were classified as school, clinic or hospital, home, work, other institution, across multiple settings, or not reported.

Therapist The individual/s responsible for implementing contingencies and or implementing treatment were classified as self, experimenter or therapist, parent, teacher or school staff, peer, staff of other agency, or not reported.

Target Behavior

Direction Goal direction of the target behavior was assessed through increasing or decreasing criterions and categorized accordingly.

Dimension The dimension/s of the target behavior being addressed was categorized as frequency, magnitude, accuracy, duration, or latency. Dimension was not recorded for CCD graphs which utilized a shaping procedure as shaping requires successive approximations of a behavior rather than a manipulation of a dimension of behavior already in an individual's repertoire.

Target Behavior A diverse range of target behaviors was expected. Therefore, specific target behaviors were recorded and classified following data collection for summation purposes. Behavior classifications included medical-related (e.g., glucose monitoring), habitual behaviors (e.g., smoking), self or home care and safety (e.g., fire safety, brushing teeth), physical and sport (e.g., walking with crutches, dart throwing accuracy), academically related (e.g., spelling, reading fluency), verbal or communicative (e.g., talking out, Braille use), compliance to requests, or other (e.g., package delivery, noise tolerance).

Pre- and Post-intervention and Intervention Characteristics

Pre-intervention Employment of pre-intervention procedures was recorded as present or absent for intervention training, functional behavioral assessment or functional analysis, and stimulus preference assessment. Any type of reinforcer assessment was incorporated including asking the individual or significant others, survey, observation, or stimuli trials. Also, use of individual or group goal-setting of criterions and or final goal was recorded.

Contingencies Each intervention was coded as having individual, group, both, or no contingency. Types of contingent reinforcement utilized were coded as token economy, tangible or activity, and or verbal praise. Punishments for not meeting

criterions were recorded and classified as time-out, response cost (e.g., monetary, token removal), overcorrection, and follow-through. Follow-through referred to forced follow-through to meet criterion; this was coded as punishment as it functions similar to overcorrection in that an avoidance response (e.g., prompt becoming a reflexive conditioned motivating operation) or behavioral fluency may develop to decrease the likelihood future criterions will not be met. Schedules of reinforcement and punishment were also noted.

Other Interventions If intervention components other than contingent reinforcement and punishment were used, these were recorded and classified as prompting, public posting or feedback, contingency contracting, self-monitoring, instruction or modeling, or other (e.g., systematic desensitization, habit reversal).

Response Opportunity Data was collected concerning response opportunity exceeding criterions and opportunity to fail.

Shaping Data was collected as to whether the CCD was used as a shaping procedure.

Changing Components Pilot data revealed articles varied in the use of the changing criterion as a level of performance expected and or level of performance given differing tasks or materials. Level of performance expected refers to direct calculations of the behavior (e.g., number of cigarettes smoked per day, percent of questions completed) whereas in cases where differing tasks or materials were present for criterions the level of performance expected did not necessarily vary. To clarify, in a study conducted by Schleien et al. (1981), two adults with severe handicaps were part of a procedure to hit the dartboard with darts. These individuals were required to throw three darts and hit the board each time in order to progress to a further distance or height closer to the standard 5 ft, 8 in. Thus, the behavioral performance of throwing three darts with 100 % accuracy was stable and the task topography of distance and height changed. Data was collected for both criterion types.

Post-intervention Presence of formal and or informal social validity was recorded as well as maintenance or follow-up data collection.

CCD Design Characteristics

Design Structure Each graph with a CCD portion was coded as an A (i.e., intervention only), AB, ABA, ABAB, or other model type. Parametric variations such as B¹, B², B³ were condensed for coding purposes into the singular B. Each graph was analyzed using a straight edge or ruler to determine whether phase length and magnitude varied between criterions at least one time or two or more instances. For example, if the

first criterion was set to 10, the second to 8, and the third to 7; phase magnitude differed at least once (two increments from criterion 1 to 2 and one increment from criterion 2 to 3). Graphs were coded as having either a full (i.e., reversal to baseline), mini (e.g., reversal to previous criterion), or no reversal. Also, it was noted if the CCD was used within a multiple baseline design and what type (i.e., across subjects, settings, or behaviors).

Baseline and Criteria In order to gauge the relationship between baseline data and the first criterion, data was collected and analyzed using the following information: the lowest point, highest point, and average performance at baseline; and the level of first criterion. This information was collected from the results section of the article or was measured using a straight edge against the graph. The total number of criteria implemented was also recorded.

Inter-rater Agreement

A second individual independently coded 15 % of articles in order to establish inter-rater agreement across coded variables. Agreement was calculated by the number of agreements on item-by-item comparison divided by the number of agreements plus disagreements and multiplied by 100. The mean inter-rater agreement was 98.2 % (range, 93.2 to 100 %) across cases.

Results and Discussion

Of the 127 references found, 19 were removed as they were book chapters or discussions that did not contain original research data, 1 article concerned animal behavior, and 1 was not written in English. The remaining 106 articles yielded a total of 267 CCD graphs (236) or intervention data (31). Articles were located in 57 different journals; see Table 2 for a summary list of journals. From these journals, 11 articles were published between 1971 and 1979 (10.4 %), 24 articles between 1980 and 1989 (22.6 %), 18 articles between 1990 and 1999 (17.0 %), 35 articles between 2000 and 2010 (33.0 %), and 18 articles between 2011 and 2013 (17.0 %).

No *range-bound* and *distributed-criterion* design data were observed, and articles describing those variations (McDougall 2005, 2006) referenced research not yet published.

Population, Setting, and Therapist

A total of 906 individuals participated in CCD research. Approximately 414 participants were children (45.7 %), 63 were adults (6.9 %), 3 studies with 169 participants had both child and adult age ranges (18.7 %), and 4 studies with 260

Table 2 Journals with published changing criterion design studies

Journal	Number of studies	Percentage of sample
<i>Journal of Applied Behavior Analysis</i>	11	10.4
<i>Education and Treatment of Children</i>	6	5.7
<i>Journal of Behavior Therapy and Experimental Psychiatry</i>	6	5.7
<i>Behavioral Interventions</i> ^a	5	4.7
<i>Behavior Modification</i>	5	4.7
<i>Education and Training in Autism and Developmental Disabilities</i>	4	3.8
<i>Journal of Behavioral Education</i>	4	3.8
<i>Journal of Developmental and Physical Disabilities</i>	4	3.8
<i>Behavior Therapy</i>	3	2.8
<i>Journal of Autism and Developmental Disorders</i>	3	2.8
<i>Research in Autism Spectrum Disorders</i>	3	2.8
<i>Augmentative and Alternative Communication</i>	2	1.9
<i>Clinical Practice in Pediatric Psychology</i>	2	1.9
<i>Journal of Organizational Behavior Management</i>	2	1.9
<i>Journal of Positive Behavior Interventions</i>	2	1.9
<i>Journal of Visual Impairment & Blindness</i>	2	1.9
<i>Psychology in the Schools</i>	2	1.9
Other journals with one publication	40	37.7

^a Formerly *Behavioral Residential Treatment*

participants did not report age (28.7 %). Of the participants, 748 had no reported diagnosis (82.6 %), 85 with a Developmental Disorder (9.4 %), 44 individuals were diagnosed with a chronic medical condition or physical disability (4.9 %), 24 individuals were classified with an emotional disorder (2.6 %), 11 with ADHD (1.2 %), 7 with LD (.8 %), and 6 with other diagnoses not fitting these categories (.7 %). Eighteen participants had comorbid conditions (2.0 %).

Of the 267 CCDs noted, the majority were conducted within schools (45.7 %), followed by clinics or hospitals (16.1 %), home (13.5 %), across the day (9.4 %), at another institution (7.5 %), at an unknown location (4.9 %), and work (3.0 %). The individual responsible for implementing the CCD was most frequently the experimenter or therapist (47.9 %), followed by a teacher or staff (27.3 %), parent (19.1 %), other staff or agency (7.5 %), unspecified (6.4 %), self (4.5 %), and a peer (2.2 %). More than one individual was involved with implementation in 40 individual CCD interventions (15.0 %).

Target Behavior

Studies utilizing shaping were excluded from direction and dimension analyses. A further study was excluded which

utilized the CCD to reduce body weight and implemented weight criterions; this study was removed as weight is not a behavior, rather a by-product of other behavior(s) and possibly confounded by biology and other factors. Of the remaining 237 studies, 152 attempted to increase a behavior (64.1 %). Frequency of behavior was the dimension most often targeted (57.4 %) as opposed to more fine manipulations such as magnitude (23.6 %) and duration (5.1 %). Within the total 267 interventions, the type of target behavior varied widely across classifications with academic-related behaviors being the most frequent (19.0 %); see Table 3 for a full list of dimension and target behavior percentages.

Pre- and Post-Intervention and Intervention Characteristics

Pre-intervention A small portion of the 267 studies conducted intervention training with participants prior to implementation (18.7 %). Functional behavior assessments (8.6 %) and preference assessments (12.4 %) were not frequently reported. A total of eight studies (4.9 %) reported participant involvement in setting criterions (i.e., goal-setting).

Contingencies Contingencies for reinforcement and punishment were most frequently individual (71.5 %) with the majority of the remaining studies implementing no type of contingency (22.8 %). A small amount of studies used group contingencies (4.5 %) or simultaneous individual and group (1.1 %). Approximately 73.4 % of interventions utilized one

or more reinforcement methods and 22.5 % utilized a punishment method; see Table 4 for percentages of reinforcement and punishment methods. No intervention used more than one form of punishment. A notable number of interventions incorporated reinforcement with another intervention (35.6 %) or used reinforcement alone as an intervention (18.0 %). Only a small amount of studies used both reinforcement and punishment without another intervention (10.1 %) and fewer studies used punishment alone (.7 %). Of the 196 interventions with contingent reinforcement, 167 (85.2 %) utilized fixed-ratio schedules (FR) and 29 used variable ratio (VR). All 61 interventions with contingent punishment implemented punishment on fixed-ratio schedules, specifically FR-1 in 58 interventions (95.1 %).

Despite a significant proportion of studies using rewards for meeting criterions, it is surprising that preference assessments were conducted in very few. When the CCD is implemented as an intervention in itself, reinforcement should be effective in shaping behavior if goal-reaching (i.e., automatic reinforcement) is not.

Other Interventions Intervention strategies other than reinforcement and punishment were implemented in approximately 70.0 % of CCD interventions. In 24.0 % of studies, an intervention that utilized neither reinforcement nor punishment contingencies was used. See Table 4 for a detailed percentage breakdown of intervention types used.

Response Opportunity The opportunity to perform beyond the criterion was available in 190 interventions (71.2 %); however, in five of these interventions participant's programmed reward was not provided if criterion was exceeded. Opportunity

Table 3 Target behavior characteristics

	237 CCD interventions		106 CCD articles	
	Frequency	Percent	Frequency	Percent
Direction				
Increase	152	64.1	62	66.7
Decrease	85	35.9	31	33.3
Dimension				
Frequency	136	57.4	49	20.7
Magnitude	56	23.6	18	7.6
Accuracy	26	11.0	10	4.2
Duration	12	5.1	10	4.2
Latency	7	3.0	6	2.5
Behavior type				
Academic	45	19.0	18	7.6
Physical/sport	38	16.0	10	4.2
Communicative	30	12.7	17	7.2
Self-care/safety	30	12.7	12	5.1
Habitual	22	9.3	8	3.4
Compliance	16	6.8	7	3.0
Medical	10	4.2	7	3.0
Other	46	19.4	14	6.0

Table 4 Contingency and intervention types

	Frequency	Percent
Reinforcement		
Tangible/activity	129	48.3
Verbal praise	90	33.7
Token economy	55	20.1
Punishment		
Response cost	22	8.2
Follow-through	19	7.1
Time-out	10	3.7
Overcorrection	9	3.4
Other components		
Feedback/public posting	72	27.0
Instruction/modeling	59	22.1
Prompting	30	11.2
Self-monitoring	15	5.6
Contingency contract	22	8.2
Other	21	7.9

beyond the criterion was restricted in 48 cases (18.0 %). In 16 cases (6.0 %), this was because the target skill had not been learned (i.e., shaping a new behavior); in 24 cases (9.0 %), procedure data for this variable was not reported. The extent of opportunity to perform beyond criterion (e.g., to final goal, to the next criterion) was not coded as almost no articles reported that detail explicitly. The opportunity to fail to meet criterion was not available in 38 interventions (14.2 %; e.g., hand over hand follow-through or materials not accessible if shaped) and not known in 8 (3.0 %). In total, 62 (23.2 %) cases restricted responding, 42 (15.7 %) excluding studies with shaping which in most cases had not taught the full skill, or changed materials so that previous and future criterions could not be performed.

Shaping The CCD was used as a shaping procedure in approximately 27 interventions (10.1 %) within 12 articles. Of these articles, four were for self-care and safety behaviors, two were related to improving physical or sport performance, two for shaping communicative responses, one for medical-related, one for academic-related, and one for other behavior.

The effective and appropriate use of the CCD with shaping procedures has been suggested in several publications (e.g., Alberto and Troutman 2006; Hall 1971; Hall and Fox 1977). When the CCD is used to shape a new behavior, the nature of the design is changed from a single topography to several and elements of control and the questions answered are altered. Shaping requires successive, topographically different approximations and therefore is not subject to some traditional elements of control utilized with the CCD: reversal and variability in phase magnitude. Also, when the CCD is implemented as a shaping procedure, success for each criterion is dependent on behaviors achieved in earlier criterions (Cooper et al. 2007). These behaviors are interdependent and graphed simultaneously; thus, the CCD does not assess responding for each approximation prior to, during, and following each criterion. This lack of appropriate documentation is avoided in other designs such as the multiple probe design (Homer and Baer 1978).

Changing Components A changing criterion of expected performance level was implemented in 236 interventions (88.4 %), the level of performance expected remained stable in 25 (9.4 %), and information was missing in 6 (2.2 %). The topography of the dependent variable, intervention, or materials varied in 71 interventions (26.6 %), remained stable in 195 (73.0 %), and information was missing for 1 study (.4 %). The majority of studies consisted of a changing criterion level of performance while the task and materials remained stable (72.7 %). A total of eight studies contributed 25 CCD data graphs that held a stable criterion of performance expected but varied the task content (9.4 %). Another example of this CCD variation involved increasing the accuracy of math problems completed by an adolescent through peer tutoring (Burley et al. 1994). The student was required to achieve 70 % or higher

accuracy for 3 days on given math problems prior to moving to the next math objective. Math objectives varied from adding fractions, subtracting fractions, and converting mixed numbers to fractions. Seventeen studies contained 39 CCD data (14.6 %) where both the performance criterion and task or material changed, of which, eight studies involved shaping. Grey et al. (2009) used this type of CCD with an intervention to increase waiting behavior in an adolescent with developmental disabilities. The intervention consisted of three phases: holding a red card up and saying “wait,” placing the red card on a timer set to the required seconds of non-tantrum behavior and saying “wait,” and having the red-wedge timer set to the required duration of appropriate behavior. The second and third phase took place in the classroom unlike the first phase, and four sessions were conducted in the cafeteria. The study lasted 122 sessions and included a total of 28 changing criterions with each criterion lasting longer than the last and for approximately four sessions. Each session successfully completed with non-occurrence of tantrum behavior was rewarded.

Variations on the CCD such as these should be interpreted with caution. Again, when the topography or materials are altered from one criterion to the next, as in shaping, experimental control is confounded and a functional relationship cannot be demonstrated (Tawney and Gast 1984). When the level of performance of a behavior, such as student completion of homework, is kept constant though the subject matter changes (e.g., social studies, math), it is important to determine whether the desired action is due to a lack of motivation or skill. In either case, intervention within another design such as a CCD within a multiple baseline design would be more advantageous in demonstrating change.

Post-intervention Of the 267 CCD data collected, social validity was collected formally or informally in 43 studies (16.1 %); all of which reported positive feedback. A maintenance or follow-up was conducted in 105 studies (39.3 %).

CCD Design Characteristics

Design Structure The majority of studies used a simple AB design (63.0 %) with a small percentage using a baseline only (4.9 %), a reversal ABAB (8.3 %), a withdrawal ABA (3.4 %), or another design (21.0 %). Approximately 70 studies combined the CCD within a multiple baseline design. A combination utilizing the multiple baseline design across behaviors was the most frequently encountered. See Table 5 for design type percentages across studies.

Just over half of the studies varied phase length between criterions at least once (55.8 %) and a considerable proportion varied length two or more times (35.6 %). Phase magnitude was varied in just under half of the studies (45.7 %) and less than half of those varied magnitude two or more times between criterions (22.1 %). Approximately 85 studies (31.8 %)

Table 5 Design structure

Design	Frequency	Percent
A	13	4.9
AB	167	62.5
ABA	9	3.4
ABAB	22	8.2
Other	56	21.0
Simple	93	34.8
Multiple baseline across		
Behaviors	48	18.0
Settings	9	3.4
Subjects	13	4.9
Repeated across		
Behaviors	5	1.9
Settings	1	.4
Subjects	93	34.9
Not reported	5	1.8
Reversal		
Mini-reversal	28	10.5
Full	28	10.5
Unknown	41	15.4
None	170	63.7

varied both phase length and magnitude once or more, 101 (37.8 %) varied either length or magnitude at least once, and 15 (5.6 %) did not vary either. Of the studies utilizing three or more criteria (183; i.e. allowing for two or more changes to occur), 27 (11.3 %) varied both length and magnitude on two or more occasions, 98 (40.8 %) varied either length or magnitude on two or more occasions, 115 (47.9 %) did not vary either twice or more, and 27 (11.3 %) had incomplete data.

Though some studies may be restricted in their ability to vary phase length or magnitude, the relatively low number of studies implementing variations was unexpected. As mentioned, the CCD is one of the least experimentally strong designs in single case research and therefore exclusion of the components that increase its validity is counterintuitive.

Approximately 63.7 % of studies conducted did not implement either a mini or full reversal. See Table 5 for reversal percentages. Again, a reversal, mini or full, is not a required design component of the CCD; however, implementation of either type of reversal considerably increases design strength by demonstrating the independent variable can increase or decrease responding accordingly. The majority of the studies did not use the CCD as a shaping procedure and therefore the dimension of the behavior being modified, in most cases, was in the individual's repertoire and should be compliant to directional shift.

Baseline and Criteria Approximately 46 interventions had zero as the highest point at baseline (17.2 %), 75 interventions

Table 6 Phase characteristics

	Frequency	Percent
Number of criteria		
1	12	4.5
2	15	5.6
3–4	84	31.5
5–6	48	18.0
7–8	14	5.2
9–10	12	4.5
11+	25	9.4
Unknown	57	21.3
Phase length variation		
0	49	18.4
1	54	20.2
2+	95	35.6
Unknown	69	25.8
Phase magnitude variation		
0	76	28.5
1	63	23.6
2+	59	22.1
Unknown	69	25.8

had incomplete data and therefore were not analyzed as to how the first criterion was chosen. Four interventions used the lowest data point at baseline as a first criterion (2.1 %), while seven used the highest point (3.6 %). Of the 181 remaining, 18 used a criterion 5 % below or above the baseline average (9.9 %). More studies started with a criterion above the mean (68.0 %) than below (28.2 %). The following percentage ranges combine criteria above and below the mean: 21 studies set the first criterion between 5.01 and 15 % (11.6 %) away from the mean of baseline, 39 between 15.01 and 35 % (21.5 %), 38 between 35.01 and 100 % (21.0 %), and 65 over 100 % (35.9 %); 48 of which had baseline averages of 0.

As previously reported, 25 interventions used a stable criterion of performance and therefore had a total of one criterion. The maximum number of criteria reported was 44. See Table 6 for criterion totals and their frequencies. As mentioned, it has been suggested that a changing criterion have at least two criterion shifts or two subphases of “B” within a study (Kazdin 2011, p. 172).

Conclusion

Results indicate that the CCD has been utilized with varied populations, in varied settings, and targeted a wide variety of behaviors; however, the majority of CCD studies were structured in ways which threatened experimental control.

Approximately 25 % (i.e., 26 studies yielded 69 graphs) of data reviewed met all requirements necessary for adequate control (see Table 1). This suggests some fundamental lack of understanding of the science underlying the CCD. Specifically, most studies lacked variation in phase magnitude (and did not compensate with a reversal), lacked variation in phase length, or implemented two or fewer changes in criterion; while some restricted responding so that criterions were forcibly met and or could not exceed the criterion. A small portion of studies utilized the CCD nontraditionally (i.e., shaped a new behavior or varied materials) which, as discussed, confounds experimental control (Tawney and Gast 1984). As demonstrated by Hall in his early articles, the best use of the CCD is in the dimensional manipulation of behaviors using static methods and behaviors across criterions.

The utility of the CCD as a behavior change agent is evident. It is a flexible design that can be used with any population, on a diverse range of behaviors, in a variety of settings, while maintaining adequate experimental control. The CCD is particularly well suited for cultivating incremental changes in behavior that may otherwise be resistant to sizable, immediate changes (e.g., habits) or aversive (e.g., anxiety-provoking). Further, it provides unique advantages over other designs such as continuous implementation of the independent variable, unlike reversal designs requiring a temporary withdrawal of intervention, and it is able to replicate intervention effect with the same individual and target behavior through manipulation of the independent variable, unlike the multiple baseline design which requires replication demonstrated across multiple persons, settings, or behaviors. These features make the CCD the ideal practitioner's design.

Clinical significance may have been demonstrated in many of the articles evaluated; however, the demonstration of the intervention as responsible for behavior change was inadequate in the majority of studies due to lack of experimental control. Certainly an increase in the fluency or rate of a functional skill, or decrease in problem behavior, whatever the method is clinically valuable; however, in a research context, lack of controlling methods does not validly inform intervention utility with specific populations or behaviors nor make valid comparisons across intervention variations. This should not, however, discourage use of the CCD within a clinical setting. Our recommendation would be to incorporate mini-reversals into the CCD whenever feasible to improve experimental control. Also, the use of forced follow through to criterion may in some cases be beneficial in encouraging independent behavior change; however, we recommend clear documentation between forced versus independent responding.

Practitioners and researchers might also consider the following when implementing the CCD, if applicable, (1) use of a functional behavior assessment or functional analysis (e.g., high rate behaviors) to create a function-based criterion

contingency when appropriate; (2) if contingent reinforcement is planned, conduct a reinforcer assessment or stimulus preference assessment to ensure the contingency has reinforcing properties to promote the likelihood of meeting criterions; (3) client/participant involvement in intervention development (e.g., goal-setting, setting contingencies) may provide added encouragement and adherence to intervention; (4) client/parent/teacher/staff implementation of contingencies to provide basic training in a behavioral management strategy for future use; and (5) a measure of social validity to examine intervention satisfaction (e.g., procedures and outcome) and social significance.

The CCD is somewhat inflexible in terms of variation due to the design features needed to achieve adequate control; however, basic variations within the classic structure which do not impact those features can be used such as the implementation of variable schedules of reinforcement or punishment. Reviewing the literature, the authors only noted three studies to implement this simple modification. Notably, De Luca and Holborn (1992) examined revolutions per minute on an exercise bicycle thru VR reinforcement within a CCD that was found to successfully increase the rate of exercise as each VR range criterion increased. De Luca and Holborn note that gains with the VR schedule inside the CCD were greater than those observed in a similar study which used FR and fixed-interval schedules within a non-CCD design (De Luca and Holborn 1990). Another intervention option that was infrequently utilized was self-management or self-reinforcement strategies such as criterion-setting, independent choosing of reward, delivering or obtaining reward independently, or self-evaluating or monitoring. Given the contingencies, the CCD appears to be fairly simple to implement. Use of self-management or self-reinforcement strategies should further reduce resources required for implementation. Further, school-based research has found these strategies to be more effective with greater generalization than teacher-managed interventions (e.g., Fantuzzo et al. 1988) where greater student involvement corresponded with larger intervention effects (Fantuzzo and Polite 1990).

One limitation of this review is that all studies that utilized a CCD structure may not have been included in analysis. This may have occurred because the search was limited for practicality to articles labeled as changing criterion design, though other studies may have been structured in a CCD format. For example, interventions that utilized progressive increments of differential reinforcement of low or high rates of behavior may have been equivalent to the CCD, but not labeled as such. Alternatively, a CCD study may not have been present on the databases used or cited in articles found.

The CCD studies examined in this review were somewhat homogenous in that the majority used individual contingencies with FR schedules of reinforcement to increase the frequency of a behavior. Future research might benefit from

examining or comparing the efficacy of the CCD with other interventions (e.g., self-monitoring) as it is an intervention in itself rather than solely a research tool. Also, future analyses exploring relationships of design components and successful behavior change may be beneficial. For example, does a specific type of initial criterion (e.g., mean of baseline, highest responding at baseline) or smaller (versus larger) changes in phase magnitude lead more frequently to adherence or success? Further, assessing skilled professional's fluency in and perceptions of the CCD may provide insight as to why it is infrequently used, though to date its prevalence in applied settings such as schools has not been examined.

References

- Alberto, P. A., & Troutman, A. C. (2006). *Applied behavior analysis for teachers* (7th ed.). Upper Saddle River: Pearson Education, Inc.
- Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: strategies for studying behavior change* (2nd ed.). New York: Pergamon Press.
- Brown, R. E., Copeland, R. E., & Hall, R. V. (1974). School phobia: effects of behavior modification treatment applied by an elementary school principal. *Child Study Journal*, 4, 125–133.
- Burley, S., Gutkin, T. B., & Wendy, N. (1994). Assessing the efficacy of an academic hearing peer tutor for a profoundly deaf student. *American Annals of the Deaf*, 139, 415–419.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River: Pearson Education, Inc.
- Delprato, D. J. (2002). Countercontrol in behavior analysis. *The Behavior Analyst*, 25, 191–200.
- De Luca, R. V., & Holborn, S. W. (1990). Effects of fixed-interval and fixed-ratio schedules of token reinforcement on exercise with obese and non-obese boys. *Psychological Record*, 40, 67–82.
- De Luca, R. V., & Holborn, S. W. (1992). Effects of a variable-ratio reinforcement schedule with changing criteria on exercise in obese and nonobese boys. *Journal of Applied Behavior Analysis*, 25, 671–679.
- Fantuzzo, J. W., & Polite, K. (1990). School-based, behavioral self-management: a review and analysis. *School Psychology Quarterly*, 5, 180–198.
- Fantuzzo, J. W., Polite, K., Cook, D. M., & Quinn, G. (1988). An evaluation of the effectiveness of teacher- vs. student-management classroom interventions. *Psychology in the Schools*, 25, 154–163.
- Flood, W. A., & Wilder, D. A. (2004). The use of differential reinforcement and fading to increase time away from a caregiver in a child with separation anxiety disorder. *Education and Treatment of Children*, 27, 1–8.
- Grey, I., Healy, O., Leader, G., & Hayes, D. (2009). Using a Time Timer to increase appropriate waiting behavior in a child with developmental disabilities. *Research in Developmental Disabilities*, 30, 359–366.
- Hall, R. V. (1971). *Managing behavior: behavior modification, the measurement of behavior*. Lawrence: H&H Enterprises.
- Hall, R. V., & Fox, R. G. (1977). Changing-criterion designs: an alternate applied behavior analysis procedure. In B. C. Etzel, J. M. LeBlanc, & D. M. Baer (Eds.), *New developments in behavioral research: theory, method, and application*. Hillsdale: Lawrence Erlbaum Associates, Inc.
- Harris, F. N., & Jenson, W. R. (1985). Comparisons of multiple-baseline across persons designs and AB designs with replication: issues and confusions. *Behavioral Assessment*, 7, 121–127.
- Hartmann, D. P., & Hall, R. V. (1976). The changing criterion design. *Journal of Applied Behavior Analysis*, 9, 527–532.
- Horner, R. D., & Baer, D. M. (1978). Multiple-probe technique: a variation on the multiple baseline design. *Journal of Applied Behavior Analysis*, 11, 189–196.
- Kazdin, A. E. (1982). *Single-case research designs: methods for clinical and applied settings*. New York: Oxford University Press.
- Kazdin, A. E. (2011). *Single-case research designs: methods for clinical and applied settings*. New York: Oxford University Press.
- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, 34, 26–38.
- McDougall, D. (2005). The range-bound changing criterion design. *Behavioral Interventions*, 20, 129–137.
- McDougall, D. (2006). The distributed criterion design. *Journal of Behavioral Education*, 15, 237–247.
- Roane, H. S., Rihgdahl, J. E., Kelley, M. E., & Glover, A. C. (2011). Single-case experimental designs. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 132–147). New York: Guilford Press.
- Schleien, S. J., Wehman, P., & Kiernan, J. (1981). Teaching leisure skill to severely handicapped adults: an age-appropriate darts game. *Journal of Applied Behavior Analysis*, 14, 513–519.
- Sidman, M. (1960). *Tactics of scientific research: evaluating experimental data in psychology*. New York: Basic Books.
- Tawney, J. W., & Gast, D. L. (1984). *Single subject research in special education*. New York: Macmillan Publishing Company.
- Wallace, C. J., & Davis, J. R. (1974). Effects of information and reinforcement on the conversational behavior of chronic psychiatric patient dyads. *Journal of Consulting and Clinical Psychology*, 42, 656–662.
- Weis, L., & Hall, R. V. (1971). Modification of cigarette smoking through avoidance of punishment. In R. V. Hall (Ed.), *Managing behavior: behavior modification applications in school and home* (pp. 54–55). Lawrence: H & H Enterprises.