Journal of Emerging Trends in Educational Research and Policy Studies (JETERAPS) 8(2): 74-89

jeteraps.scholarlinkresearch.com

A Systematic Review of Classroom Interventions to Promote Social Inclusion for Preschoolers with Special Needs In Inclusive Educational Settings

Dr. Vicky Tsang and Libby Cheng
Education University of Hong Kong
Tai Po, Hong Kong

Abstract
Three dimensions of social inclusion are identified: social relationships, social acceptance status and social interaction. Ten intervention studies in social inclusion for preschoolers with special needs between 2006 and 2016 were sought from 4 databases, ProQuest, Science Direct, Academic Search Premier, and PsycINFO and reviewed using Reichow's criteria for identifying evidence-based practices (EBP) and the newly-developed ecological validity indicators. The review investigated:(i) the construct validity, (ii) the utility validity, (iii) the EBP ratings and (iv) the ecological validity of the reviewed studies. All reviewed studies focused on evaluating the effect of intervention on improving social interaction skills of the preschoolers with special needs. The effect of intervention was primarily assessed from a child-centered intra-personal perspective. EBP assessment yielded 2 strong, 4 adequate and 4 weak studies. Ecological validity showed 100% studies scored 'full' in social interaction; 20% scored 'partial' in social acceptance; and none (0%) measured social relationship. Past social inclusion intervention studies investigated primarily from the dimension of social interactions. Furthermore, interventions targeted primarily at children with special needs but not their counterparts, such as their typically developing peers or teachers. Our findings calls for an adoption of inter-personal assessment measures across multiple stakeholders and longitudinal studies across multiple measures. This shift to an ecological framework will have a significant impact on the future development of social inclusion practice as it implies every stakeholder in our community of diversity needs to learn how to interact with one another to bring about true inclusion.

Keywords: social inclusion; evidence-based practice; ecological validity; preschool; systematic review

INTRODUCTION

The initiative of inclusive education has transformed school systems around the world to include children with special needs in regular schools and preschools since the International Conference on inclusive education held in Geneva in 2008 (UNESCO, 2008). Three indicators of inclusive education have been put forward to assess the success of its development, namely, (i) the physical presence of children with special needs within the regular schooling system, (ii) their full and active participation within school life, and (iii) academic achievement which demonstrates their optimal learning capabilities (Farrell, Squires, & Armstrong, 2011).In real life practice, it is relatively easier to ensure the physical presence of children with special needs in school settings by using school enrolment statistics and to demonstrate their achievement gain from evidence on academic performance. However, there remains little empirical evidence to show that children with special needs physically included in regular school settings have also been socially accepted (Kasari, Locke, Gulsrud, &Rotheram-Fuller, 2011). For example, children with autism can learn to sit still throughout the entire class period, but may not be engaged in a learning activity that is consistent with the rest of the class. Furthermore, past social inclusion studies addressing peer relationship and social acceptance mainly targeted school population (Chamberlain, Kasari, &Rotheram-Fuller, 2007; Frostad, &Pijl, 2007; Siperstein, Glick, & Parker, 2009), little was known for the pre-schoolers learning in integrated kindergarten settings. So whether preschoolers with special needs are actually socially included in the classroom remains a hypothesis yet to be confirmed. To what extent have past social inclusion intervention studies achieved their aims? This review study aims to provide some answers to these questions using a systematic scientific approach.

LITERATURE REVIEW

Construct Validity Indicators for Social Inclusion

Past research in inclusive education found that the core underlying motive of parents to send their preschoolers with special needs into regular schools was the increased opportunities of social interaction (Bossaert, Colpin, Pijl, &Petry, 2013). A common assumption by the parents is that the more prolonged and extensive social contacts made with their peers, the more positive effect will be on their social-emotional development (Koster, Nakken, Pijl, &Van Houten, 2009). Apparently, social inclusion is an important agenda for inclusive education from many stakeholders' views. According to Cullinan, Sabornie and Crossland (1992), social inclusion implies: (i) being accepted as a member of a group; (ii) have at

least one mutual friendship; and, (iii) participate socially in equivalent status in group activities. Therefore, social inclusion is best reflected by the extent of friendship and social network established based on the interpersonal relationship of children with and without special needs, as well as the extent of social acceptance in the social setting. Furthermore, the meta-analysis by Koster's team (2009) on social inclusion intervention studies also supported the subthemes of social inclusion but adding the social interaction component: (i) social relationships between children with and without special needs; (ii) social acceptance status in class; and (iii) social interactions between children with and without special needs.

However, it is more sophisticated to foster social inclusion than to promote social interaction skills. The former is an inter-personal construct involving the changes between at least two persons or social parties; in contrast to the latter which is an intra-personal concept related to child characteristics. As such, an intervention that shows its effectiveness in promoting social interaction skills, such as increased initiation, and response to name or question in inclusive settings, does not necessarily lead to enhancement in social inclusion, such as improved reciprocal friendship and social participation. From the socio-ecological perspective, human development cannot adequately be understood by merely focusing on individual functioning within each system. Rather, the ecological environment should be conceived topologically as a nested inter-related arrangement of structures, each contained within the next (Bronfenbrenner, 1977; Shaffer&Kipp, 2010; Wu & David, 2002). It is imperative to study the reciprocal social influences between different child-related human systems in the social network, such as the family system, the school system and the community system. Therefore, it is essential to develop a social-ecological perspective for the construct of social inclusion.

Utility Validity Indicators for Social Inclusion

Having a valid social inclusion measure is important to evaluate the effectiveness of social inclusion intervention. However, assessing inter-personal relationship appears more sophisticated than assessing intra-personal characteristics since perspectives from multiple stakeholders are required in the former. Paststudies attempted to use rating scales and inventories to assess social status for children with special needs in schools (Bauminger&Kasari,2000; Hunt, Soto, Maier, &Doering, 2004). Until the last decade, there have been growing attempts in social inclusion studies to adopt sociometric techniques, such as sociograms or peer nomination tests, which provide a context for assessing inter-personal peer relationships (Mikami et al., 2013; White, Keonig, &Scahill, 2007).

Evidence-Based Practice Indicators for Social Inclusion

Past studies on social inclusion from the perspective of social acceptance status yielded inconsistent findings. Some found that children with special needs are less popular, have fewer friendships and are less likely to be a member of a group as compared to their typically developing peers(Freeman &Alkin, 2000; Kasari, Locke, Gulsrud, &Rotheram-Fuller, 2011; Guralnick, Connor, Hammond, Gottman, &Kinnish, 1996). Other studies found that a majority of children with special needs do have one or more friends and do have a reasonable number of interactions with peers in the classroom (Koster, Pijl, Nakken, & Van Houten, 2010; Pijl,Frostad, &Flem, 2008). In order to help preschoolers with special needs socially accepted by and make reciprocal friendships among their peers of typical development in the inclusive classroom, it is important to adopt reliable and valid Evidence-Based Practices (EBP) criteria for assessing social inclusion interventions. Reichow's team (2011) put forward five EBP criteria to evaluate effectiveness of intervention studies or strategies, including: (i) the evaluation of both single case and group research within the same review, (ii) rubrics with operational definitions, (iii) the delineation of primary (essential) and secondary (non-essential) quality indicators, (iv) guidelines for the determination of research report strength (e.g., strong, moderate, weak) and (v) criteria for the overall determination of an EBP.

STUDY OBJECTIVES

As such, the objectives of this systematic review were to: (i) explore the construct validity of social inclusion in the reviewed studies; (ii) investigate the utility validity of the measures used to evaluate social inclusion; (iii) evaluate the evidence-based practice (EBP) ratings of social inclusion interventions in reviewed studies; and, (iv) assess the ecological validity of the reviewed interventions using our newly developed indicators of social inclusion.

METHOD

Search Procedures

The authors of this review examined the existing literature on social inclusion intervention studies for preschoolers with special needs inregular preschool settings. In order to determine effective social inclusion practices for preschoolers with special needs, the authors conducted a systematic review using combinations of the following terms: "classroom", "inclus*", "preschool* or kindergarten", "special need* or disabilit*", "intervention or program or training", "assessment or measure", and, "social or emotion" in four search e-databases, namely, ProQuest, Science Direct, Academic Search Premier, and PsycINFO. The search was restricted to English language peer-reviewed studies published between 2006 through

2016. This timeframe was chosen in order to include studies not covered in past reviews prior to that period and to highlight the increasing contributions of recent research to this field. In addition to the electronic search, the authors also completed an ancestral search of all articles uncovered by hand searching in the reference lists of relevant articles related to preschoolers with special needs and social inclusion intervention.

Selection Criteria

Four eligibility criteria were adopted for the selection of review articles, including studies that: (i) used a single case experimental design that had sample size with less than twenty participants by comparing result across subjects, behaviors or settings, or group research design that had sample size with twenty or more participants (Knight &Sartini, 2015). (ii) recruited participants with special needs between the age of 3-7 years; (iii)adopted interventions that were designed to promote social interaction, and/or social network and peer acceptance for preschoolers with special needs; and,(iv) implemented in inclusive preschool settings, that is, having children with and without special needs in the same classroom.

Search Procedures

All search outputs were independently examined by two reviewers to determine eligibility for inclusion. The researchers rated each article either positive (+) indicating that the article "adequately satisfies all the criteria", or negative (-) for those that "do not adequately satisfy all of the criteria", or indeterminate (?) indicating "information not sufficient to be judged". Using the search keywords, the titles and abstracts were first screened to identify relevant articles. Full texts were further reviewed using established evaluation criteria (i.e., Tables 1, 2 and 3) for those abstracts with positive or indeterminate ratings.

Critical Appraisal and Assessment Procedures

Construct assessments. For the first and second objectives regarding the construct and utility validity of social inclusion and related assessment measures, we adopted a qualitative analysis approach. First,we attempted to identify the most commonly used themes for measuring social inclusion. Then, from the study findings we synthesized the themes into several dimensions of social inclusion. The results were used for discussing what the best construct indicators and valid measures for the construct of social inclusion would be for assessing the effectiveness of related intervention studies. Based on the findings, the ecological validity of the intervention studies was explored.

Evidence-based practice and utility validity assessments. For the third objective regarding evaluation of the EBP status of the interventions and the embedded comprehensive strategies, the reviewed studies were first assessed according to Reichow's six primary quality indicators (QIs), including: (i) participant characteristics (P) consisting of age and gender (P1), operationalized diagnosis (P2), interventionist characteristic (P3), standardized test scores (P4); (ii) independent variables (IV) that classified as either IV with replicable precision definition (IV1), or IV defined without specific details (IV2), or insufficiently defined IV (IV3); and (iii) dependent variables (DV) that subdivides into operational precision (DV1), details for replicating the measure (DV2), measure linking to DV (DV3), and data collected during intervention (DV4); (iv) baseline conditions (BL) that encompass at least 3 measurement points (BL1), or show stable visual analysis (BL2), or show no trend indication (BL3), or provide details for replicating the baseline (BL4); (v) visual analysis (VA) that shows stable level or trend (VA1), contains less than 25% overlap of data points between adjacent conditions (VA2), shows a large shift in level or trend between adjacent conditions (VA3); and, (vi) experimental control (EC). For each of these categories with its independent rating criteria, the rater graded the quality strength of the study and allocated each a rigor rating of either high (H), acceptable (A), or unacceptable (U).

Furthermore, the methodological quality of the interventions that used both single design (SCED) and group research design (GROUP) can be assessed using Reichow's formula (Knight &Sartini, 2015):
$(GROUP_S*30) + (GROUP_A*15) + (SCED_S*4) + (SCED_A*2) = Z$

In order to determine the EBP status, Z score for each intervention strategy can be calculated. $GROUP_S$ is the number group research design studies rated strong; $GROUP_A$ is the number group research design studies rated adequate; $SCED_S$ is the number of participants for whom the intervention strategy was applied to a SCED studies rated strong; $SCED_A$ is the number of participants for whom the intervention strategy was applied to a SCED studies rated adequate; and, Z is the total number of points for an intervention (p. 32, Reichow, 2011). Based on the range of the resulting Z score, the selected practices were further classified as: (E) established EBP when Z score is equal to 60, (P) probable EBP when Z score is between 31–59, or,(NE) not an EBP when Z score is between 0–30.

Furthermore, to evaluate theEBP status of the comprehensive intervention strategies, the rater graded the established EBP interventions with an EBP rating, either as 'strong' if the study contained 70%or more EBP rating; as 'medium' if the study contained 31-69% EBP rating; or, as 'weak' if the study contained less than 30% EBP rating.

In addition to the six primary QIs ratings, the rater also noted the presence (+) or absence (-) of the six secondary QIs with its independent rating criteria, including: (i) inter-observer agreement (IOA) with more than 80% reliability across all conditions, raters, and participants; (ii) kappa (K) with score > 60% for at least 20% of sessions across all conditions; (iii) blind raters (BR) if raters are blinded to the treatment condition of the participants; (iv) fidelity (F) with measurement statistics > 80%; (v) generalization or maintenance (G/M) if outcome measures are collected after the final data collection to assess G/M; and, (vi) social validity (SV) that subdivides into socially important DVs (SV1), time- and cost-effective intervention (SV2), comparisons between individuals with and without disabilities (SV3), clinically significant behavioural change (SV4), consumers satisfied with results (SV5), independent variables manipulated by familiar contact person to participant (SV6), conducted in a natural context (SV7) with containing at least 4 SVs.

Consequently, the rater assigned the overall rigor rating for each study as being strong (S) if it received high ratings on all primary QIs and at least three of the secondary Qis; adequate (A) if it received high ratings on four or more primary QIs and at least two secondary Qis; or, weak (W) when it received fewer than four high ratings on primary QIs or less than two secondary QIs.

Ecological validity assessments. For the forth objective, the authors proposed three social inclusion indicators to evaluate the ecological validity of the 10 social inclusion intervention studies under review, namely, (i) social relationships, (ii) social interactions, and (iii) social acceptance status. The rater assigned the rating on each indicator as 'full' if the measurement encompassed the mutual or reciprocal responses from 'the target stakeholder group' (e.g., children with special needs or parents) and 'the counterparts of the target stakeholder group' (e.g., children without special needs or teachers); 'partial' if the measurement encompassed the responses from either 'the target stakeholder group' or 'the counterparts of the target stakeholder group'; and 'none' if the measurement did not encompass the responses from any stakeholders.

RESULTS

Data Extraction

The keyword search found 689 publications, including Science Direct; Academic search premier; PsycINFO and Proquest. After excluding2 duplicates, the title review further excluded 612 articles that did not meet the inclusion criteria. The subsequent abstract review on the remaining 75 articles further excluded 52 publications, out of which 35 were found not related to intervention, 14 not related to preschoolers and 3 not carried out in inclusive educational settings. Consequently, 23 articles were extracted for full article review. Among them, a total of 13 articles were further excluded because six studies were not related to intervention; five studies were not for preschoolers and two studies were not implemented in an inclusive educational setting Figure 1, (See Appendix). Furthermore, none of the GROUP studies and 10 SCED studies met the inclusion criteria. Therefore, only 10 SCED studies were rated in this review.

Interrater Reliability

Interrater reliability was conducted after the second author coded each study according to Reichow (2011) indicators. Two post-graduate researchers conducted reliability testing on four randomly assigned articles coded for quality criteria (40 % of the articles) and two of the articles coded for descriptive information (30 % of the articles). Using a point-by-point method, the co-author divided the number of agreements by the total number of indicators, and then multiplied by 100. The researchers obtained an acceptable reliability score of 83.3 % for quality criteria and 82 % for descriptive information. Most disagreements were related to participants, baseline, and visual analysis.

Study Quality Of Reviewed Interventions

EBP ratings for quality analysis of reviewed studies

The ratings of all primary and secondary QIs and the overall rigor ratings for each selected study is reported in Table 1(See Appendix).

The EBP ratings and the frequency of intervention strategies in all reviewed studies are reported in Table 2. (See Appendix).

Out of the 10 reviewed studies, two achieved strong EBP ratings (i.e., #3 and #5),fourwere rated adequate (i.e., #4, #6, #7, and #8),and, four studies scored weak (i.e., #1, #2, #9, and #10) (Table 1). Furthermore, established EBP strategies were found in five of the ten intervention studies (i.e., #3, #4, #5, #7, and #9). Among them, four studies (i.e., #3, #4, #5, and #7) contained 100% EBP rated strategies (Table 2).

A total of nine types of intervention strategies were identified, namely, Peer-mediation, Positive-feedback, Visual-support, Response-prompting, Incidental-teaching, Audio-support, Prompt-fading, Token economy system, and Extinction. However, only studies that were rated strong and adequate (i.e., #3, #4, #5, #6, #7, and #8) were assessed for EBP status. Z score for each type of strategy was calculated using Reichow's formula, only four of the nine strategies (i.e., Peer-mediation, Positive-feedback, Visual-supports, Response-prompting) obtained a Z score above 60 and can be considered as strategies with established EBP (E). Another five

strategies (i.e., Incidental teaching, Audio support, Prompt fading, Token economy system, and Extinction) obtained a Z score between 0-30 and were considered as not an EBP (NE).

Among the four EBP rated strategies, Positive-feedback and Response-prompting were the most commonly used, adopted in nine of the 10 interventions (90%). Specifically, in Positive-feedback, eight of the nine studies (89%) used positive reinforcement (i.e., #1, #2, #3, #4, #5, #7, #8, and #9) and three studies (33%) used verbal feedback during intervention (i.e., #2, #6, and #7). In Response-prompting**,** seven of the nine studies (78%) used verbal prompts (i.e., #1, #2, #3, #4, #5, #7, and #8); five studies (56 %) used physical prompts (i.e., #2, #5, #7, #9, and #10), three studies (33%) used gestural prompts for correct response (i.e., #5, #7, and #8). In Peer-mediation, four of the 10 studies (40%) used peer modeling during intervention (i.e., #3, #4, #5, and #7). In Visual-supports, three of the seven studies (43%) used picture icon showing communication strategies (i.e., #5, #6, and #7), two of the seven studies (29%) used scripts cards and activity schedule or written schedule (i.e., #9 and #10), one of the seven studies (14%) used visual cue (i.e., #3), or, communication board (i.e., #4) (Table 2).

Utility Validity of Reviewed Studies

The utility validity data of the reviewed interventions were analyzed based on the descriptive quality analysis of the reviewed interventions (Table 3). (See Appendix).

Participants. The reviewers examined the participant characteristics of all 10 reviewed studies according to the four primary quality indicators of an EBP, namely, age and gender, operationalized diagnosis, interventionist characteristics, and standardized test scores (Reichow, 2011). All 10 reviewed studies (100%) reported the age and gender and interventionist characteristics for its participants. Only three studies (30%) provided information of the operationalized diagnosis for its participants (i.e., #3, #4, and #5). Six of the 10 studies (60%) provided information of the standardized test scores for its participants, such as the BDI cognition scores, PLS-3, PLS-4 scores, IQ scores, and VABS scores (i.e., #3, #4, #5, #6, #7, and #8). As a result, only three of the 10 studies (30%) were rated high (H)as each contained all four participant characteristics (i.e., #3, #4, and #5). Another three studies (30%) were rated adequate (A)as each only contained three of the four participant characteristics (i.e., #6, #7, and #8). The remaining four (40%) were rated unacceptable (U) (i.e., #1, #2, #9, and #10) because they failed to provide participants information on either age and gender, interventionist characteristics, or standardized test scores. It is worthy to note that five of the 10studies (50%) have involved teachers as the intervention implementers (i.e., #2, #3, #4, #7, and #10). The remaining five studies (50%)involved researchers, educational interventionist or therapist as the intervention implementer(i.e., #1, #5, #6, #8, and #9) (Table 3).

Independent variables. Nine of the 10 studies (90%) had number of social initiation to peer as independent variables (i.e., #1, #2, #3, #5, #6, #7, #8, #9, and #10); four of them (40%) also had response to the requester or speaker(i.e., #3, #7, #8, and #9); three (30%) had engagement duration (i.e., #4, #5, and #7) as independent variables (Table 3).

Dependent variables. Only study #4 (10%) recorded on the interaction duration of participant with peers alone. Nine of the 10 studies (90%) recorded on the number of initiation and response to peer for participants as their dependent variables (i.e., #1, #2, #3, #5, #6, #7, #8, #9, and #10), two studies(20%) also recorded the duration of engagement during play time(i.e., #5 and #7) (Table 3).

Inter-observer agreement. Nine of the 10 reviewed studies (90%) accomplished the acceptable criteria of at least 80%for inter-observer agreement (IOA) (i.e., #1, #3, #4, #5, #6, #7, #8, #9, and #10). Only one study (10%) had below 80% IOA and therefore rated non-acceptable (Table 3).

Kappa.Only two of the 10 reviewed studies (20%) provided their kappa (K) value (i.e., #6 and #7). All the remaining eight studies (80%) did not provide information about their K values (Table 3).

Blind raters.Two of the 10 reviewed studies (20%) were blind rated to their treatment conditionsand social validity results (i.e., #6 and #7). Two otherstudies (20%) were non-blind rated to their treatment conditions but blind-rated for the social validity results (i.e., #3 and #4). All the remaining six studies (60%) were non-blind rated to their treatment conditions and social validity results (Table 3).

Fidelity. Seven of the 10 reviewed studies (70%) included measures of procedural integrity with fidelity scores ranged from 83 % to 100% (i.e., #1, #3, #4, #5, #6, #7, and #8). The remaining three studies did not present any data on procedural (Table 3).

Generalization. Four studies (40%) showed generalization of target social behavior for the participants (i.e., #2, #3, #8, and #9). Among them, two studies reported generalization across untrained peers and settings for all participants (i.e., #3 and #8); whereas another two studies found that participants generalized the on-task behavior in another setting (i.e., #2). One study demonstrated response generalization in novel unscripted responses (i.e., #9) (Table 3).

Maintenance. The participants in seven studies found maintenance effect of the targeted social behavior for 1 to 10 months (i.e., #1, #2, #3, #4, #5, #8, and #9) (Table 3).

Social validity. Six of the 10 reviewed studies (60%) included social validity measures (i.e., #1, #3, #4, #5, #6, and #7). Of these, one used Likert scales (i.e., #1), another used a questionnaire (i.e., #3), four used Likert scales questionnaires (i.e., #4, #5, #6, and #7), and one of them also used checklist and follow up questions (i.e., #3). None of the studies (0%) used interviews. All these studies reported positive results in social validity. As for example, teachers and teaching staff found the training used in one intervention practical and simple, they also rated the intervention as significant and effective to the training of children with autism (i.e., #5). Four of the six(67%) studies found that teachers were satisfied with the intervention and considered them socially important to the participants (i.e., #3, #4, #6, and #7) (Table 3).

Outcomes. All 10 studies in this review demonstrated positive intervention outcomes. Four of these studies (40%) found that the participants increased in social initiation (i.e., #1, #2, #5, and #10); four studies in the number of social responses during peer interaction (i.e., #3, #6, #7, and #8); one study in the engagement duration during play session (i.e., #5); yet another one in the duration ofextended peer initiations (#4), the child's response to peer initiations (i.e., #9), the child's commenting behavior and verbal or non-verbal request (i.e., #6) and teachers' attention to students with special needs (i.e., #2) after intervention (Table 3).

Multiple stakeholders measures. All 10 studies (100%) reported child to peer's reciprocal responses, only two of these studies (20%) also reported teacher to student's reciprocal responses(i.e., #1 and #2). None of the 10 reviewed studies (0%) reported parent to child's reciprocal responses (Table 3). (See Appendix).

Ecological validity of reviewed social inclusion interventions

For the three proposed social inclusion dimensions to evaluate the ecological validity, namely, Social relationships, Social interactions and Social acceptance status, all of the reviewed studies (100%) scored 'Full' in the rating of Social interaction**.** Specifically, two of these studies (20%) measured the mutual responses from two stakeholder groups during social interaction, that is, between student with special needs and peers with or without specials needs, and, student with special needs and teacher in the inclusive classroom (i.e., #1 and #2) while the remaining eight studies recorded the mutual responses between student with special needs and peer with or without specials needs only.

It is worth-noting that two of the 10 reviewed studies (20%) included measurement of Social acceptance status and therefore was scored with partial ratings in its ecological validity in social inclusion (i.e., #7 and #10). One study (i.e., #7) provided child or peer rated positive and negative response to social response initiation as an indicator of peer acceptance and rejection, the other study (i.e., #10) provided peer rated attitude scale scores (Table 3). None of the 10 studies (0%) measured Social relationship between any two stakeholder groups.

DISCUSSION

In terms of construct validity, the 10 reviewed social inclusion intervention studies primarily examined the dimension of social interactions between children with and without special needs in the inclusive classroom. This can be seen from their set independent and dependent variables which focus on social interactions, as for instances, the number of social initiation to peer, the responses to the requester, and the interaction engagement duration. The emphasis in social interaction is also reflected by the measures used to evaluate social inclusion in the selected intervention studies (i.e., #3measured the frequency of initiation and responses to low-probability requests by using Event Recording Scheme; #5measured forms and duration of communication by using Personal Digital Assistant-based data collection system; whereas #6 & #7measured verbal or non-verbal request and responses by using Peer Language and Behavior Code. All these measures targeted at evaluating the effect of intervention on improving the social interaction skills of the preschoolers with special needs as perceived by teachers and/or students.

Such a focus on the dimension of social interaction in the reviewed intervention studies suggested that the social interaction skills training that the children with special needs received would help them to be more socially included in the regular classroom. However, these measures in the reviewed studies only assessed the intra-personal changes of social interaction performance in children with special needs. Only when the inter-personal social interaction transactions are also recorded by appropriate measures can we be sure of the effectiveness of the social inclusion interventions. Therefore, we recommend the social interaction assessment of multi-stakeholders in the inclusive settings. That is, not only do we measure the changes of social interaction in children with special needs, we also assess the changes of responsiveness from their peers and teachers or caregivers in such social interaction transactions. In recent studies, more socio-metric assessment measures, such as the sociograms, have been used to evaluate the changes in reciprocal social interaction between children with and without special needs (Kuhne& Wiener, 2000; Roberts &Zubrick, 1992;

Stone & La Greca, 1990). None of our reviewed studies provided sociometric data on the assessment of social relationship except two of the 10 studies (i.e., #7 and #10), however, these studies were not rated as high evident-based practice. Therefore, although it seems that the utility validity of studies might increase by using the sociometric measures, the validity and reliability of such measuring methods still await more evidence-based testings.

Furthermore, the inter-personal dimension of social interaction inspires us to rethink about the design of our social inclusion intervention programs. It is because not only do children with special needs need to be trained to interact with their peers but also their counterparts of typically developing peers need to learn how to interact with those who do not initiate, do not possess prosocial behaviours and even at times display aggressive or problematic social behaviours in the classroom. Social inclusion intervention, which often focuses in the training of social interaction skills needs to be carried out in a reciprocal manner. That is, it should involve the reciprocal interactions between children with and without special needs, and, between teachers and their students with special needs.

For the dimension of social acceptance status, only two of the 10 reviewed studies adopted measures to assess social acceptance for preschoolers with special needs, namely, positive and negative response to other's social initiation (i.e., #7), and, attitude scale that was completed by typically developing peers(i.e., #10).Even so, only post intervention data was collected. Both studies did not make a comparison of the pre and post intervention of peer acceptance, neither did they measure from an inter-personal reciprocal friendship perspective. Therefore, they did not score full ecological validity rating on social acceptance status. The downplay in the dimension of social acceptance in social inclusion intervention is a forceful alarm for social inclusion advocates as it reveals the missing gap in our current social inclusion curriculum. It confronts us with the critical question: Have our current preschool curriculum include that of social inclusion? If yes, is the curriculum implementation really achieving the social validity that it aims to achieve? Our findings found as low as 40% of studies earned the EBP social validity scores.

In addition, only questionnaires or rating scales to solicit one stakeholders' views were adopted from participants' parents and/or teachers, as for instances, questionnaires were filled by parents and teachers (i.e., #3 & #5). From the ecological validity point of view, the social acceptance status would need to be assessed from an inter-personal reciprocal manner as well. Furthermore, it is worth-noting that not only the uni-directional nominating or being nominated, but also the bi-directional nomination need to be counted as an indicator of reciprocal social relationship building (Frostad&Pijl, 2007; Pijl, Frostad, &Flem, 2008; Nabuzoka, 2003).

For the dimension of social relationships, the most commonly used EBP strategies identified from the EBP assessment, namely, Positive-feedback and Response-prompting, have informed interventionists, such as teachers, to make use of the EBP strategies to mediate classroom environment or routine so as to bring about social inclusion in the classroom. Irvin, Boyd & Odom (2015) assessed the specific type of teacher feedback, Adult-talk, in inclusive preschool settings and found that teachers tended to focus more on using their feedback to ensure children participate in classroom activities. Furthermore, it was found that children with different special needs received different types of Adult talk. Children with more severe autistic characteristics and atypical behaviours were more likely to receive more Adult talk related to behavioural management than those for supporting engagement in peer relations. In contrast, teachers tended to offer more Adult talk for those children who exhibit more self-initiated interactions (Pelletier et al.,2002; Reeve, 2013). Hence, rather than merely adopting a one-way instructional approach and providing teacher-directed step-by-step prompting for children with special needs in classroom learning tasks, teachers should also attempt to use response-prompting in learning situations whereby positive social contact and peer relations can be fostered in classroom (Irvinet al.,2015). Moreover, the response-prompting strategies should not only be used on preschoolers with special needs, but also on their typically developing peers, such that they will learn how to interact appropriately with children with special needs in the inclusive classroom.

Last but not least, the low proportion of reviewed studies (i.e., 30%) that yielded generalization and maintenance effect calls for more longitudinal intervention studies across multiple measures over long periods of time (Locke et al., 2012). The failing long-term effect might be accounted by the lack of learning contexts to practice the learnt skills (Grizenko et al., 2000; Soresi& Nota, 2000). Future social inclusion design can focus on creating learning contexts for promoting social relationship building and nurturing social acceptance climate.

LIMITATON OF STUDY

Similar to the systematic review findings which also adopted Reichow (2011) EBP criteria by Knight and Sartini (2015), no group research design studies were found in our study that met criteria for review. In addition, the Reichow's EBP criteria was originally used in both single-case and group design research studies involving individuals with autism spectrum disorders, its validity in assessing studies in other special needs conditions is yet to be tested.

Moreover, in contrast to Reichow's review which primarily evaluated research reports of intervention with single strategy, such as video modeling, or peer-mediated intervention, all of the studies in our review used multiple strategies in their intervention. The authors in our study could only identify the EBP status of the strategies that were used in each intervention, but were unable to identify the EBP status of the intervention with multiple strategies using Reichow's EBP ratings.

Furthermore, since none of the effective strategies identified in the review were used in isolation, future research should conduct component analyses to determine the most beneficial aspects of the intervention strategy (Knight and Sartini, 2015). As for example, while the strategy of Response-prompting was rated as effective, which aspect, how and where to implement the strategy is crucial to determine its fidelity. Therefore, as field experts continue to explore the evaluative criteria needed to determine an EBP, emphasizing on the methodological fidelity, creating effective learning contexts or climate in applied settings may worth-considering as well. It also points to the importance of social validity where feedback from stakeholders can inform us on the core determinants that lead to its effectiveness in practice.

ACKNOWLEDGEMENTS

I acknowledge that the wider study from which this paper is generated is financially supported by the General Research Fund under University Grants Council of Hong Kong Special Administration Region, China [Grant number: GRF 18405414].

REFERENCES

Bauminger, N., &Kasari, C. (2000). Loneliness and friendship in high-functioning children with autism. *Child Development, 71,* 447-456. Doi:10.1111/1467-8624.00156

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32*(7), 513-531. doi:10.1037/0003-066X.32.7.513

Bossaert, G., Colpin, H., Pijl, S. J., &Petry, K. (2013). Truly included? A literature study focusing on the social dimension of inclusion in education. *International Journal of Inclusive Education, 17*(1), 60-79. doi:10.1080/13603116.2011.580464

Bursuck, W. D. (1983). Sociometric status, behavior ratings, and social knowledge of learning disabled and low-achieving students. *Learning Disability Quarterly, 6*(3), 329-338. doi:10.2307/1510444

Chamberlain, B. A., Kasari, C. A., &Rotheram-Fuller, E. (2007). Involvement or isolation? The social networks of children with autism in regular classrooms. *Journal of autism and developmental disorders. 37,* 230–242. doi: 10.1007/s10803-006-0164-4

Cullinan, D., Sabornie, E. J. &Crossland, C. L. (1992) Social mainstreaming of mildly handicapped students. *Elementary School Journal, 92*(3), 339-351. doi:10.1086/461696

Farrell, P., Squires, G., & Armstrong, D. (2011). Inclusive education for children with special educational needs: current uncertainties and future directions. In *Children with Special Educational needs – Considering the Whole Child.* (pp. 79-98). London: Routledge. Publication link: 71903914-0564-4f4b-80df-773ed3a71ba1

Freeman, S. F. N., &Alkin, M. (2000). Academic and social attainments of children with mental retardation in general education and special education settings. *Remedial and Special Education,*21(1), 3–18. doi:10.1177/074193250002100102

Frostad, P., &Pijl, S. J. (2007). Does being friendly help in making friends? The relation between the social position and social skills of pupils with special needs in mainstream education. *European Journal of Special Needs Education,22*(1), 15-30. doi:10.1080/08856250601082224

Grizenko, N., Zappitelli, M., Langevin, J., Hrychko, S., El-Messidi, A., Kaminester, D., Pawliuk, N. and TerStepanian, M.2000. Effectiveness of a social skills training program using self/other perspective-taking: a nine-month follow-up. *American Journal of Orthopsychiatry*, 70(4): 501–509. doi:10.1037/h0087662

Guralnick, M. J., Connor, R. T., Hammond, M., Gottman, J. M., &Kinnish, K. (1996). Immediate effects of mainstreamed settings on the social interactions and social integration of preschool children. *American Journal on Mental Retardation*, *100*(4), 359-377. doi: http://psycnet.apa.org/psycinfo/1996-01602-003

Hunt, P., Soto, G., Maier, J. &Doering, K. (2004). Collaborative teaming to support students at risk and students with severe disabilities in general education classrooms. *Council for Exceptional Children,* 69(3), 315–332. doi:10.1177/001440290306900304

Irvin, D. W., Boyd, B. A., & Odom, S. L. (2015). Child and setting characteristics affecting the adult talk directed at preschoolers with autism spectrum disorder in the inclusive classroom. *Autism, 19*(2), 223-234. doi:10.1177/1362361313517398

Kasari, C., Locke, J., Gulsrud, A., &Rotheram-Fuller, E. (2011). Social networks and friendships at school: Comparing children with and without ASD. *Journal of autism and developmental disorders*, *41*(5), 533-544. doi:10.1007/s10803-010-1076-x

Knight, V. F., &Sartini, E. (2015). A comprehensive literature review of comprehension strategies in core content areas for students with autism spectrum disorder. *Journal of autism and developmental disorders*, *45*(5), 1213-1229. doi:10.1007/s10803-014-2280-x

Koster, M., Nakken, H., Pijl, S. J., &Van Houten, E. J. (2009). Being part of the peer group: A literature study focusing on the social dimension of inclusion in education. *International Journal of Inclusive Education*, *13*(2), 117-140. doi:10.1080/13603110701284680

Koster, M., Pijl, S. J., Nakken, H., & Van Houten, E. (2010). Social participation of students with special needs in regular primary education in the Netherlands. *International Journal of Disability, Development and Education*, *57*(1), 59-75. doi:10.1080/10349120903537905

Kuhne, M., & Wiener, J. (2000). Stability of social status of children with and without learning disabilities. *Learning Disability Quarterly*, *23*(1), 64-75. doi:10.2307/1511100

Locke, J. J., Kasari, C., Rotheram-Fuller, E., &Kretsmann, M. (2012). Among elementary school-aged children with and without an autism spectrum disorder. *School Mental Health*, 5, 38 - 47. doi: 10.1007/s12310-012-9092-y

Mikami, A. Y., Griggs, M. S., Lerner, M. D., Emeh, C. C., Reuland, M. M., Jack, A., &Anthony, M. R. (2013). A randomized trial of a classroom intervention to increase peers' social inclusion of children with attention-deficit/hyperactivity disorder. *Journal of consulting and clinical psychology*, *81*(1), 100. doi:http://doi.org/10.1037/a0029654

Nabuzoka, D. (2003). Teacher ratings and peer nominations of bullying and other behaviour of children with and without learning difficulties. *Educational Psychology*, *3*(3), 307-321. doi:10.1080/0144341032000060147

Pelletier, L. G., Séguin-Lévesque, C., &Legault, L. (2002). Pressure from above and pressure from below as determinants of teachers' motivation and teaching behaviors. *Journal of educational psychology*, *94*(1), 186. doi:10.1037/0022-0663.94.1.186

Pijl, S. J., Frostad, P., &Flem, A. (2008). The social position of pupils with special needs in regular schools. *Scandinavian Journal of Educational Research*, 52(4), 387-405. doi:10.1080/00313830802184558

Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of Educational Psychology*, *105*(3), 579. doi:10.1037/a0032690

Reichow, B. (2011). Development, procedures, and application of the evaluative method for determining evidence-based practices in autism. In B.

Reichow et al. (Eds.). *Evidence-Based Practices and Treatments for Children with Autism* (pp. 25-39). doi:10.1007/978-1-4419-6975-0_2

Roberts, C., &Zubrick, S. (1992). Factors influencing the social status of children with mild academic disabilities in regular classrooms. *Exceptional children*, *59*(3), 192-202.doi:10.1177/001440299305900303

Shaffer, D.R., &Kipp, K. (2010). *Developmental psychology: childhood and adolescence.* (8th ed). Belmont, Calif.: Wadsworth Cengage Learning.

Siperstein, G. N., Glick, G.C., & Parker, R. G. (2009). Social inclusion of children with intellectual disabilities in a recreational setting. *Intellectual and Developmental Disabilities*, 47 (2), 97-107. doi: 10.1352/1934-9556-47.2.97

Soresi, S., &Nota, L. (2000). A social skill training for persons with Down's syndrome. *European Psychologist*, 5(1): 34–43. doi:10.1027//1016-9040.5.1.34

Stone, W. L., & La Greca, A. M. (1990). The social status of children with learning disabilities: A reexamination. *Journal of learning disabilities*, *23*(1), 32-37.doi:10.1177/002221949002300109

UNESCO (2008), *Final Report of 48th session of the International Conference on*

Education: Inclusive education: the way of future. Paris: UNESCO.

White, S. W., Keonig, K., &Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of autism and developmental disorders*, *37*(10), 1858-1868. doi:10.1007/s10803-006-0320-x

Wu, J., & David, J. L. (2002). A spatially explicit hierarchical approach to modeling complex ecological systems: Theory and applications. *Ecological Modelling*, *153*(1), 7-26. doi:10.1016/S0304-3800(01)00499-9

APPENDIX

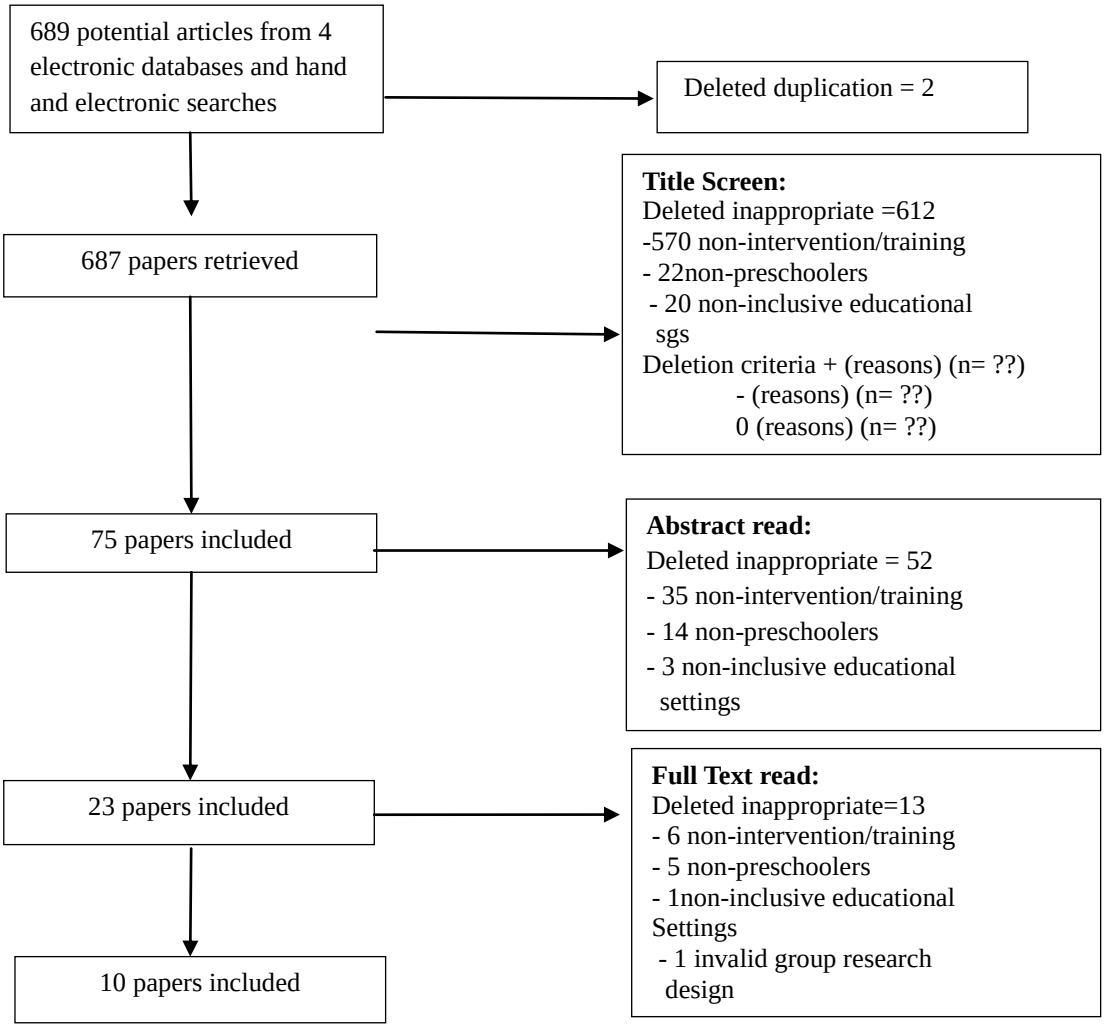

Figure 1. Flow chart of search strategy and study selection

Table 1 The overall rigor ratings of reviewed studies on social inclusion intervention

	Reviewed Study	Primary Quality Indicators							Secondary Quality Indicators					Overall Rigor Rating:
		P:	IV:	DV:	BL:	VA:	EC:	IOA:	K:	BR:	F:	G/M:	SV:	
#1	Chan and O'Reilly, (2008)	U	A	H	A	A	H	+	-	-	+	+	+	W
#2	Hundert, (2007)	U	A	H	A	A	A	-	-	-	-	+	+	W
#3	Jung et al., (2008)	H	H	H	H	H	H	+	-	-	+	+	+	S
#4	Katz and Girolametto, (2013)	H	H	H	A	H	H	+	-	-	+	+	+	A
#5	Nelson et al., (2007)	H	H	H	H	H	H	+	-	-	+	+	+	S
#6	Stanton-Chapman and Brown, (2015)	A	H	H	A	H	H	+	+	+	+	-	+	A
#7	Stanton-Chapman and Snell, (2011)	A	H	H	A	H	H	+	+	+	+	+	+	A
#8	Tzanakaki et al., (2014)	A	H	H	H	H	A	+	-	-	+	+	-	A
#9	Wichnick et al., (2010)	U	H	H	H	H	H	+	-	-	-	+	-	W
#10	Woods and Poulson, (2006)	U	A	H	A	H	H	+	-	-	-	-	-	W

Abbreviations:
P=Participants included P1-Age and Gender, P2-Operationalized diagnosis, P3-Interventionist characteristics, and P4-Standardized Test Scores with rating criteria: H contains P1-4; A contains P1, P3 and P4; U does not meet the above criteria;
IV=Independent Variables with rating criteria: H contains IVs defined with precise details; A contains IVs defined without specific details; U contains IVs defined not sufficiently;
DV=Dependent Variables included DV1-Define with operational precision, DV2-Provides details for replicating the measure, DV3-Measure links to DV, and DV4-Data is collected during intervention with rating criteria: H contains all 4 DVs; A contains 3 DVs; U contains less than 3 DVs;
BL=Baseline includedBL1-Encompass at least 3 measurement points, BL2-Show stable visual analysis, BL3-No trend indication, BL4-Provide details for replicating the baseline with rating criteria: H contains all 4 BLs; A contains 3 BLs; U contains less than 3 BLs;
VA=Visual Analysis included VA1-Show stable level or trend, VA2-Contain less than 25% overlap of data points between adjacent conditions, VA3-Show a large shift in level or trend between adjacent conditions with rating criteria: H contains 100% of all 3 VAs; A contains >66% of 2 VAs; U contains <66% of 2 VAs;
EC=Experimental Control with rating criteria: H contains at least 3 demonstrations of the experimental effect; A contains 2 demonstrations; U contains 1 demonstration;
IOA=Inter-observer agreement with IOA > 80%; K=Kappa with score > 60% for at least 20% of sessions across all conditions; BR=Blind Raters if raters are blinded to the treatment condition of the participants; F=Fidelity with measurement statistics > 80%; G/M=Generalization/ Maintenance if outcome measures are collected after the final data collection to assess G/M; SV=Social validity included SV1-Socially important DVs, SV2-Time- and cost-effective intervention, SV3-cAmparisons between individuals with and without disabilities, SV4-clinically significant behavioral change, SV5-Consumers satisfied with results, SV6-independent variables manipulated by familiar contact person to participant, SV7-conducted in a natural context with rating criteria: + contains at least 4 SVs; - contains less than 4 SVs;
+ = present; - = absent; S=Strong: received high ratings on all primary quality indicators (QIs) and at least 3 of the secondary QIs are showed; A =Adequate: received high rating on 4 or more primary Qis,no unacceptable rating on any primary QIs and at least 2 secondary QIs; W=Weak: received high ratings on fewer than 4 primary QIs or less than 2 secondary QIs;.

List of reviewed studies:

#1 = Chan, J. M., & O'Reilly, M. F. (2008). A social stories(tm) intervention package for students with autism in inclusive classroom settings. *Journal of Applied Behavior Analysis, 41*(3), 405-409. doi:http://doi.org/10.1901/jaba.2008.41-405
#2 = Hundert, J. P. (2007). Training classroom and resource preschool teachers to develop inclusive class interventions for children with disabilities: Generalization to new intervention targets. *Journal of Positive Behavior Interventions, 9*(3), 159-173. doi:http://journals.sagepub.com/doi/abs/10.1177/10983007070090030401
#3 = Jung, S., Sainato, D. M., & Davis, C. A. (2008). Using high-probability request sequences to increase social interactions in young children with autism. *Journal of Early Intervention, 30*(3), 163-187. doi:http://journals.sagepub.com/doi/abs/10.1177/1053815108317970
#4 = Katz, E., & Girolametto, L. (2013). Peer-mediated intervention for preschoolers with ASD implemented in early childhood education settings. *Topics in Early Childhood Special Education, 33*(3), 133-143. doi:http://journals.sagepub.com/doi/10.1177/0271121413484972

#5 = Nelson, C., McDonnell, A. P., Johnston, S. S., Crompton, A., & Nelson, A. R. (2007). Keys to play: A strategy to increase the social interactions of young children with autism and their typically developing peers. *Education and Training in Developmental Disabilities, 42*(2), 165-181. doi:http://www.jstor.org/stable/23879993
#6 = Stanton-Chapman, T., & Brown, T. S. (2015). Facilitating commenting and requesting skills in 3-year-old children with disabilities. *Journal of Early Intervention, 37*(2), 103-118. doi:http://doi.org/10.1177/1053815115598005
#7 = Stanton-Chapman, T. L., & Snell, M. E. (2011). Promoting turn-taking skills in preschool children with disabilities: The effects of a peer-based social communication intervention. *Early Childhood Research Quarterly, 26*(3), 303-319. doi:http://doi.org/10.1016/j.ecresq.2010.11.002
#8 = Tzanakaki, P., Grindle, C. F., Dungait, S., Hulson-Jones, A., Saville, M., Hughes, J. C., & Hastings, R. P. (2014). Use of a tactile prompt to increase social initiations in children with autism. *Research in Autism Spectrum Disorders, 8*(6), 726-736. doi:http://doi.org/10.1016/j.rasd.2014.03.016
#9 = Wichnick, A. M., Vener, S. M., Pyrtek, M., & Poulson, C. L. (2010). The effect of a script-fading procedure on responses to peer initiations among young children with autism. *Research in Autism Spectrum Disorders, 4*(2), 290-299. doi:http://doi.org/10.1016/j.rasd.2009.09.016
#10 = Woods, J., & Poulson, C. L. (2006). The use of scripts to increase the verbal initiations of children with developmental disabilities to typically developing peers. *Education & Treatment of Children, 29*(3), 437-457. doi:http://www.jstor.org/stable/42899895

	Type of strategy Reviewed / Study: Type of intervention	Peer mediation	Positive Feedback	Visual Support	Response Prompting	Incidental Teaching	Audio Support	Prompt Fading	Token Economy System	Extinction
EBP rating(only included studies that are rated as strong or adequate for Z score calculation)										
	Z score	68	88	78	76	16	0	30	10	12
	EBP Status of Strategy	E	E	E	E	NE	NE	NE	NE	NE
Descriptive information of intervention strategies for each reviewed study										
(Strong EBP rating studies)										
#3	**Jung et al., (2008):** High-Probability Request Sequences	Peer models with the high-p request sequences	Positive reinforcement during intervention	Visual cue for implementer	Verbal prompts to redirect from inappropriate behavior	-	-	-	-	Used for non-responding to request
#5	**Nelson et al., (2007):** Keys to Play Intervention	Peer models with instructor's direction	Positive reinforcement during intervention	'Key to Play' visual strategy	Indirect/direct verbal or gestural prompts, and least to most physical prompts	Play activities	-	-	-	-
(Adequate EBP rating studies)										
#4	**Katz and Girolametto, (2013):** Peer-Mediated Social Intervention	Peer models applying the communication strategies	-Positive reinforcement for child's initiation and appropriate response	-Illustrated communication Board -Storybook about developing friendship	Verbal prompt to initiate or maintain conversation	-	-	-	-	-
#6	**Stanton-Chapman and Brown, (2015):** Social Communication Intervention	-	-Verbal feedback during review	-Picture icon for communication -Computer generated story book for dramatic play	-	-	-	-	-	-
#7	**Stanton-Chapman and Snell, (2011):** Peer-based Social Communication Intervention	Peer models with instructor's direction	-Positive reinforcement -Verbal feedback during review	-Picture icon for communication, -Computer generated story book for dramatic play	Least to most verbal, gestural, and physical prompts	-	-	Systematic fading of Prompts	-	-
#8	**Tzanakaki et al., (2014): Tactile Prompt Intervention**	-	-Positive reinforcement	-	Tactile, verbal, and gestural prompts for redirection	-	-	Systematic fading of prompts	Use of tokens	-
(Weak EBP rating studies)										
#1	**Chan and O'Reilly, (2008):** Social Stories Intervention Package	-	-Positive reinforcement	-	Verbal prompts for correction	-	-	-	-	-
#2	**Hundert, (2007):** Teacher Developed Social Inclusion Intervention	-	-Positive reinforcement, -Teacher'sfeedback	-	Physical and verbal prompts	-	-	-	-	-
#9	**Wichnick et al., (2010):** Script-Fading Procedure	-	-Positive reinforcement	-Activity schedules -Written prompts	Most to least physical prompts	-	Pre-recorded scripts	Physical prompts andfading of script	Use of tokens	-
#10	**Woods and Poulson, (2006):** Use of Scripts	-	-	-Scripts cards -Written schedule	Most to least physical prompts	-	Language Master audio-taped cards	Systematic fading of manual guidance	-	-
	Frequency of use in intervention	4	9	7	9	1	2	4	2	1

Table 3 Descriptive analysis of reviewed studies on social inclusion intervention using EBP quality and ecological validity indicators

	EBP validity / Reviewed Study: Type of intervention	Primary Quality Indicators			Secondary Quality Indicators			Ecological validity Indicators*		
		Participants (P)	Independent Variables (IVs)	Dependent Variables (DVs)	Inter-observer agreement (IOA) Kappa (K) / Blind Raters (BR) Fidelity (F) / Generalization (G) Maintenance (M) / Social Validity (SV)	Outcomes (measures)	Multiple Stakeholders	Social Relationships	Social Interaction	Social Acceptance status
(Strong rating studies)										
#3	**Jung et al. (2008):** High-probability request sequences	P1: 3 males aged 5-6 P2: Diagnosed as ASD, PDD-NOS with CARS P3: Teacher or experimenter P4: BDI -- Cognition: 30-43 months; Receptive Language: 23-32 months; Expressive Language: 16-30 months	IV-1: High-Probability request sequences IV-2: Unprompted and prompted initiation and subsequent responses request	DV-1: % of correct responses to low-probability requests DV-2: # of social interaction (i.e., social initiation and responses)	Mean IOA: 98-99% K: no information BR: Non-blind rated to treatment condition but blind-rated for SV result F:100% IOA for procedural integrity checklist G: the increased IV-3 were generalized to untrained peers and novel setting. M: the increased DV-3.1 and DV-3.2 were maintained without the prompts. SV: The questionnaire of the participants' parents and teachers rated the intervention goals are socially important to the participants.	All three participants' Increased in DV-1, DV-2 and DV-3.2; Decreased in DV-4 with intervention. (DVs were measured by 6-s partial interval and event recording scheme.)	Student-peer: DV-1, DV-2, DV-3.1, DV-3.2 Teacher-student: not reported Parent-student: not reported	None	Full Responses to peer's request (DV-1, DV-2) Social initiation and responses to peers (DV-3)	None
#5	**Nelson et al. (2007):** Keys to Play intervention	P1: 4 males aged 4-5 P2: Diagnosed as mild to moderate or severe ASD with ADOS and/or CARS and or GARS P3: Research assistant P4: PLS-3 for participant 3 are 3 standard deviations below the norm	IV-1: Play initiation IV-2: Use Key to Play or strategy to enter a playgroup IV-3: Engagement time in a playgroup	DV-1: % of correct and successful initiations of play DV-2: % of using communication form (e.g., verbal, gesture, position, or object play) used during play DV-3: % of time in each play category	Mean IOA: 96-99% K: no information BR: Non-blind rated to treatment condition F: Mean Fidelity: 96% G: no information M: DV-1 maintained up to 4 weeks for all participants. SV: 70% of the teaching staff and parents rated the intervention is significant and effective. 90% of the teaching staff indicated that the intervention is easy to manage and will use it.	All four participants' Increased in DV-1 and DV-3 with intervention. (DVs were measured by a Personal Digital Assistant-based data collection system.)	Student-peer: DV-1, DV-2, DV-3 Teacher-student: not reported Parent-student: not reported	None	Full Play initiation (DV-1) Length of play with peers (DV-3)	None
(Adequate rating studies)										
#4	**Katz and Girolametto (2013):** Peer-Mediated Social Intervention	P1: 2 males and 1 female aged 4.1-5.1 P2: Diagnosed as ASD with ADOS scored 31-33 &CARS scored 98-111	IV-1: Social skills training sessions IV-2: Play session with peers	DV-1: # of extended interactive engagement DV-2: Average length of the extended interactions	IOA: 89.9% with joint interaction IOA: 95% without joint interaction K: no information BR: Non-blind rated to treatment condition but blind-rated for SV result F: 100% G: no information	All three participants' Increased in DV-1, DV-2 with intervention. (DVs were measured by interval coding system.)	Student-peer: DV-1, DV-2 Teacher-student: not reported Parent-student: not reported	None	Full Interactions duration with peers (DV-1, DV-2)	None

		P3: Early childhood educator and researcher P4: IQ scored 98-111, PLS-4 scored 57-102 VABS scored 69-110, PLS-4 scored 95-125			M: DVs maintained up to 4 to 5 weeks. SV: The questionnaire rated by the educators indicated that the intervention goals are socially important to the participants.					
#6	**Stanton-Chapman & Brown (2015):** Social communication intervention	P1: 3 males and 3 females aged 3-4 P2: Diagnosed as developm ent or language delay without operationa lized diagnosis P3: Interventio nist with education degree and min. of 5-year teaching experienc e P4: PLS-4 scored 50-89, CTRF scored 40-61, SSRS scored 60-89	IV-1: Dramatic play themes	DV-1.1: # of child's commentin g behaviors DV-1.2: # of child's requests for verbal behaviors DV-1.3: # of child's requests for non-verbal behaviors DV1-4: # of child's non-verbal requests	IOA: 84-100% K: 0.82-0.94 BR: Blind rated to treatment condition and SV result F: 93-100% G/M: no information SV: All teachers rated the behavior changes as being socially important and the intervention procedures are socially acceptable, mean rate ranged from 34-40.	All six participants' DVs increased with intervention. (DVs were measured by Peer Language and Behavior Code.)	Student-peer: DV-1, DV-1.2, DV-1.3, DV-1.4 Teacher-student: not reported Parent-student: not reported	None	Full Comment on peer's activities or make verbal/non-verbal requests to peer (DV-1.1) Initiation to peer (DV-1.2)	None
#7	**Stanton-Chapman & Snell (2011):** Peer-based social communication intervention	P1: 9 males and 1 female aged 4-5 P2: Diagnosed as developm ent or language delay or problem behavior without operationa lized diagnosis P3: Teacher P4: PLS-4 scored 60-100, CTRF scored 10-73, SSRS scored 72-98	IV-1: Dramatic play themes per Reciproc ity dyad	DV-1.1: # of turn-taking per Reciprocity DV-1.2: Duration of the child engageme nt in play with peers DV-1.3: # of the initiations and responses with an immediate peer response per Reciprocity DV-1.4: # of child/peer positive and negative response to an initiation	IOA: 86-97% K: 0.79-0.91 BR: Blind rated to treatment condition F: 95% G/M: no information SV: The rating range of the acceptability of intervention procedures was 38-40 (range from 0-40). For dyads A, B, C, and D, the mean ratings range of baseline video clips on social importance was 11.5-17 (range from 0-19), and the mean ratings range of intervention clips on social importance was 32.5-39.5 (range from 30-40).	The social communication intervention was highly effective for five children; but moderately effective for three children, and mildly effective for two children by their increase in DV-1,2, and 3. (DVs were measured by Peer Language and Behavior Code, duration measure and event measure. Play session were transcribed by using Systematic Analysis of Language Transcripts software)	Student-peer: DV-1.1, DV-1.2, DV-1.3, DV-1.4 Teacher-student: not reported Parent-student: not reported	None	Full Turn-taking with peer (DV-1.1) Interaction duration with peers (DV-1.2) Initiations and responses with peer (DV-1.3)	Partial Child/peer positive and negative response to an initiation (DV-1.4)
#8	**Tzanakaki et al (2014):** Tactile prompt intervention with and without Systematic Fading of the Prompt (SFP)	P1: 3 males aged 4-7 (study without SFP), and 1 male and 1 female aged 7-9 (study with	IV-1: Child's initiation with tactile prompt IV-2: Child's initiation with	DV-1: # of child's initiation DV-2:# of peers' responses to target child's initiation	IOA: 80-100% (study without SFP), 96-100% (study with SFP) K: no information BR: Non-blind rated to treatment condition	All three participants'DV-1 increased with intervention. (DVs were measured by frequency count during observation.)	Student-peer: DV-1, DV-2, Teacher-student: not reported Parent-student: not reported	None	Full Initiations and peers' responses (DV-1, DV-2)	None

		SFP) P2: Diagnosed as ASD without operationalized diagnosis P3: Therapist P4: IQ (SB-IV) scored 83-102, VABS' composite score is 68-79, communication score is 74-93, socialization score is 68-80, and daily living skills is 75-83	fading the strength of the tactile prompts IV-3: Child's initiation with reinforcement fading procedure		F: 97-100% (study without SFP), no data (study with SFP) G: participants were trained to generalize responses to different intervention setting, people, and delayed delivery of reinforcement M: DVs maintained during the 6 weeks follow up SV: No measurement for both studies					
(Weak rating studies)										
#1	**Chan & O'Reilly (2008):** Social stories intervention package	P1: 2 males aged 5-6 P2: Diagnosed as ASD without operationalized diagnosis P3: Researcher P4: no data	IV-1: Make social Initiation appropriately IV-2: Do not make monosyllables or noises during lesson IV-3: Make relevant comments to classroom activities IV-4: Raise hand vertically and above the shoulder	DV-1: # of inappropriate social interaction DV-2: # of inappropriate vocalizations DV-3: # of social interaction DV-4: % of opportunities of raising hand	IOA: 88-100 % K: no information BR: Non-blind rated to treatment condition F: 96% (ranged 83-100 %) G: no information M: DVs maintained up to 10 months. SV: Teachers, teaching assistants and parents rated for the importance of skills taught, perceived effectiveness, appropriateness, and future use of the intervention, which ranged from 3 to 5 (mean rating 4.3) from a Likert scale.	All participants increased in DV-1,3, and 4; but decreased in DV-2. (DVs were measured by frequency count during observation)	Student-peer: DV-1, DV-3, Teacher-student: DV-4 Parent-student: not reported	None	Full Social interaction with peer (DV-1, DV-3) Students' opportunities of raising hand in group setting with teacher and peers (DV-4)	None
#2	**Hundert (2007):** Teacher developed social inclusion intervention	P1: 5 males and 3 females' students aged 3.1-5.5, and 4 females' teachers with 7-22 years teaching experience P2: Diagnosed as ASD, Moderate developmental delay, Cerebral palsy, language delay without operationalized diagnosis P3: Teacher for	IV-1: Teacher pay attention to children in inclusive group IV-2: Peer interaction program IV-3: On-task Behavior program	DV-1: % of teacher focus on inclusive groups DV-2: % of interactive play DV-3: % of on-task behavior	IOA: 75-100% K: no information BR: Non-blind rated to treatment condition F: no information G: slight to moderate increases in child DVs (on-task behavior during circle time). M: DVs retained after 3 months' follow-up SV: not measure	Associated increase in teacher interactions (DV-1) with inclusive groups of children; mixed results for child target behaviors (DV-2 and DV-3) during training (DV-1 was measured by Eco-Behavioral System for Complex Assessments of Pre-school Environment, DV-2 and DV-3 were measured by child behavior codes)	Student-peer: DV-2 Teacher-student: DV-1, Parent-student: not reported	None	Full Interactive play with peer (DV-2) Teacher's focus to student (DV-1)	None

		students and supervisor for teacher P4: no data								
#9	**Wichnick et al. (2010):** Script- fading procedure	P1: 2 males and 1 female aged 5-7 P2: Diagnosed as ASD without operationalized diagnosis P3: Researcher P4: no data	IV-1: Independent response to peer initiation IV-3: Script-fading procedure IV-3: Novel responses to peer initiations	DV-1: # of response to peer initiation DV-2.1: # of scripted responses to peer initiations DV-2.2: # of unscripted responses to peer initiations DV-3: # of novel responses to peer initiation	mean IOA: 98-100% K: no information BR: Non-blind rated to treatment condition F: no information G: the novel unscripted responses to peer initiation demonstrated response generalization. M: DV-2 maintained at a high level after the script-fading procedure SV: not measure	All three participants' DVs increased with intervention. (DVs were measured by frequency count and verbatim recording device for responses.)	Student-peer: DV-1, DV-2.1, DV-2.2, DV-3 Teacher-student: not reported Parent-student: not reported	None	Full Responses to peer initiations (DV-1, DV-2.1, DV-2.2, & DV-3)	None
#10	**Woods and Poulson (2006):** Use of Scripts	P1: 3 males aged 5-6 and 8 grade 2 typically developing peers P2: Diagnosed as ASD without operationalized diagnosis P3: Teacher P4: no data	VI-1: Verbal initiation to peers by using scripts	DV-1: # of scripted initiations to peers DV-2: Attitude scale scores (filled by TD peers)	Mean IOA: 86-97% K: no information BR: Non-blind rated to treatment condition F: no information G/M: no information SV: not measure	All three participants' DV-1 increased with intervention; DV-2 scores showed a greater peer acceptance of participants with disabilities. (DV-1 was measured by frequency count, DV-2 was measured by the Lower Elementary Level Acceptance Scale)	Student-peer: DV-1, Teacher-student: not reported Parent-student: not reported	None	Full Initiation to peers (DV-1)	Partial Peer's attitude (DV-2)

#1-10 refer to the list of reviewed studies in table 1;P1=age and gender, P2=operationalized diagnosis, P3=interventionist characteristics, P4=standardized test scores.
*Ecological validity indicators=included measurementsthat encompass responses from:(i) the target stakeholder group (e.g., children with special needs or parents), (ii) the counterpartsof (i) (e.g., children with typical development or teachers), and (iii) the mutual / reciprocal responses from both groups (i &ii).
'Full' if (iii) is reported; 'Partial' if only (i) or (ii) is reported; 'None'if none of the stakeholders (i) to (iii) is reported.