Assessing Affective Dimensions of Play in Psychodynamic Child Psychotherapy via Text Analysis

Sibel Halfon[1]*, Eda Aydın Oktay[2], and Albert Ali Salah[2]

[1] Bilgi University, Department of Psychology,
Istanbul, Turkey
`sibel.halfon@bilgi.edu.tr`
[2] Boğaziçi University, Department of Computer Engineering,
Istanbul, Turkey
`{eda.aydin,salah}@boun.edu.tr`

Abstract. Assessment of emotional expressions of young children during clinical work is an important, yet arduous task. Especially in natural play scenarios, there are not many constraints on the behavior of the children, and the expression palette is rich. There are many approaches developed for the automatic analysis of affect, particularly from facial expressions, paralinguistic features of the voice, as well as from the myriads of non-verbal signals emitted during interactions. In this work, we describe a tool that analyzes verbal interactions of children during play therapy. Our approach uses natural language processing techniques and tailors a generic affect analysis framework to the psychotherapy domain, automatically annotating spoken sentences on valence and arousal dimensions. We work with Turkish texts, for which there are far less natural language processing resources than English, and our approach illustrates how to rapidly develop such a system for non-English languages. We evaluate our approach with longitudinal psychotherapy data, collected and annotated over a one year period, and show that our system produces good results in line with professional clinicians' assessments.

Keywords: Play therapy, affect analysis, psychotheraphy, natural language processing, Turkish language, Valence, Arousal

1 Introduction

Clinical work with young children often relies on emotional expression and integration through symbolic play [58]. Play naturally provides a venue in which children can communicate and re-enact real or imagined experiences that are emotionally meaningful to them [23, 52]. Many child therapists use play therapy to help children express their feelings, modulate affect, and resolve conflicts [16].

---

* This is uncorrected author proof. Please cite this paper as Sibel Halfon, Eda Aydin Oktay, Albert Ali Salah, "Assessing Affective Dimensions of Play in Psychodynamic Child Psychotherapy via Text Analysis," HBU 2016: 15-34.

Affective analysis of psychodynamic play therapy sessions is a meticulous process, which requires many passes over the collected data to annotate different aspects of play behavior, and the markers of affective displays. Both the verbal and non-verbal content of the interactions contain valuable information, and are analyzed in detail. Recent developments in multimedia analysis suggest that automatic tools could be used to help the analyst in these tasks. The advantages are many; such tools can support the therapist with immediate and rich feedback about the data, highlighting promising patterns for which more effort can be devoted, and also provide additional quantification of treatment effects. The disadvantages are that good automatic systems typically require a large amount of data for training, their generalization abilities may suffer from factors that may appear trivial to the experimenter (e.g. amount of ambient light, if a camera-based system is employed), and depending on the model used, justification of the classifications may be difficult to fathom.

In this work, we propose such an automatic, text-based tool for affective content analysis from verbal communications of children during play activity in psychodynamic treatment. Automatic analysis of psychodynamic play therapy is not a broadly researched subject, and we hope that our contribution will initiate more research in this domain. Another important point is that our tool is based on the Turkish language, which is spoken by more than 70 million people worldwide, but for which few analysis tools are available[3]. We make the developed tool available to the research community.

1.1   Preliminary Research Questions

It will be useful to put the work presented in this paper into the broader context of our research program. Using a naturalistic process-outcome design of psychodynamic play therapy with children at an outpatient clinic, our experimental study assessed affect expression over the course of treatment using two different kinds of instruments. Children's Play Therapy Instrument is a psychodynamically informed measure that aims to assess the structure and narrative of a child's play activity in psychotherapy [31]. The affective dimensions of the measure allows the rater to code an array of emotions expressed by the child while playing. The second instrument we use is the automated affective analysis model for Turkish language that analyzes affect from text using dimensions of Valence and Arousal [4]. Children's natural linguistic output over the course of treatment is assessed with the use of this instrument, and it is this second instrument that we describe in detail in this paper.

Given the paucity of research with clinical children in treatment, we report here a preliminary study which aims to investigate the utility of using an automatic text analysis tool to study the relations between affective expression in psychodynamic play therapy as it relates to different types of psychopathology and coping and its changes over the course of treatment. In terms of the

---

[3] Ethnologue estimates it as 71 millions as of 2006, related Wikipedia content suggests the numbers to be closer to 80 millions.

type and quality of affective expression in play, literature shows that children with behavioral problems are likely to express more negative affect. However, there have been very few studies that looked at these associations with clinical samples in therapy. The first aim of this study was to investigate the relations between the type and quality of affect expressed in play and its relation to type of psychopathology. Literature shows that different negative emotions relate differently to Internalizing and Externalizing Problem behaviors. In general, irritability and anger has been hypothesized to predict Externalizing Problem behaviors, whereas sadness, anxiety, and fear are believed to predict Internalizing Problems (see [20] for a review). Therefore, in our research, we specifically look at Internalizing and Externalizing children's expression of anger, sadness and anxiety in the initial stages of treatment, as well as over the course of treatment.

Secondly, studies show that the expression of negative affect in play is related to better coping in the long-run [53]. Play provides a context in which a child is able to explore both positive and negative emotional content in a safe, controlled manner. Play ultimately provides the opportunity to increase positive affect and reduce negative affect. However, empirical evidence to support this theory with clinical children over the course of treatment is limited. The second aim of this study was to assess the type of affect expressed in play over the course of psychodynamic play therapy and its relation to different kinds of psychopathological functioning.

Based on literature, several specific hypotheses can be tested for the initial phase of psychotherapy and over the course of treatment. The first hypothesis is that children with Externalizing Problems will show higher levels of anger and lower levels of valence. The second hypothesis is that children with Internalizing Problems will show higher levels of sadness, anxiety and lower levels of valence. Finally, we hypothesize that in the initial phase of therapy, both Internalizing and Externalizing children are expected to bring more negative affect (high anger, sadness and low valence) followed by more positive affect (high valence) over the course of treatment.

The two assessment instruments mentioned earlier, one used by psychologists, the second introduced in this paper, both aim to quantify affect over the course of the therapy for the investigation of these hypotheses.

The paper is structured as follows. In Section 2 we summarize related work in the area of affective expression in play. We broadly describe affect in psychotherapy research, specifically discuss the role of text analysis, and then briefly overview text analysis for sentiment and affect detection, which is a widely researched topic for multimedia and information retrieval. Section 3 introduces our text analysis system. Section 4 describes the data, and the participants of the study. Section 5 reports our experimental results, including sensitivity analysis for parameters of the system and ablation study for measuring the contribution of the different parts of the system. Finally, Section 6 concludes the paper.

2   Related Work

2.1   Affect in psychotherapy research with children

Affect plays a significant role in psychotherapy, and a model of emotions can be used to explain different aspects of psychopathology [48]. In psychotherapy, the emphasis is on the analysis of affect rather than the elicitation of particular emotions, as the latter is quite difficult. Play therapy is one approach to obtain rich behavioral data with affective content.

There are numerous studies that link children's behavior in play to affective states. Children with disruptive behaviors have been shown to display more negative affect in their play and lower levels of affect regulation [17, 59, 11]. Dunn and Hughes found that children who were hyperactive and displayed conduct problems showed more physical aggression in their pretend games [19]. Similarly, children with disruptive behavior disorders such as Conduct Disorder and Attention Deficit Hyperactivity Disorder show more hostility and anger in their play [14]. Von Klitzing et al. found that expressing negative and/or aggressive affect in disorganized pretend play predicted behavior problems [63].

Russ & Cooperberg found that first and second graders who had more negative affect in their early play also had more symptoms of depression when measured 10 years later [51]. Additionally, in a sample of 322 six year-olds, some of whom were exposed to cocaine prior to birth, negative affect in play significantly correlated with both Internalizing and Externalizing behaviors [57]. Negative affect in play also correlated significantly with Major Depression Disorder and Oppositional Defiant Disorder in this study. These studies point to the importance of the relation between negative affect in play and behavioral problems. Some studies have also looked at the longitudinal effects of expression of affect, especially negative affect in play and behavioral functioning. Marcelo and Yates evaluated prospective relations among preschoolers' pretend play, coping flexibility, and behavior problems across varied degrees of child stress exposure [35]. They found that preschoolers who expressed more negative affect in their play engaged in more varied coping strategies (i.e., coping flexibility) during a simultaneous delay of gratification challenge and fewer Internalizing Problems one year later. These results show that even though expression of negative affect may initially be related to higher frequency of behavior problems, it may be related to enhanced coping in the long run [54].

However, there is a gap between the research literature that shows that affect in play facilitates coping, and the actual process of what happens in play therapy with clinical children in terms of affective changes. Bratton, Rhine & Jones, in a meta-analysis of outcome of play therapy, identified only seven studies that reported that play overall helped in the reduction of anxiety and fear [7]. The few empirical studies in the play intervention area that were focused on play with specific problems found that play reduced fears and anxiety for children with an acute physical illness and separation anxiety [5, 41, 47]. The research findings from a variety of studies in the child and adult areas suggest that other types of negative affect, like anger should also be helped by play therapies however these

studies have not been carried out. There is even less research about the kinds of affective transformations that take place over the course of treatment. Gaensbauer and Siegel found that children who expressed affect in play, especially negative affect, were better able to work through their trauma in play-based therapy [26]. According to them, the key element that enables a child to use play adaptively, is the "degree to which the affects can be brought to the surface so the child can identify them and integrate them in more adaptive ways" (p. 297). Singer proposed that children can then increase positive affect and reduce negative affect through play [61]. This conceptualization fits with the idea that play is one way in which children learn to regulate their emotions. However, these ideas need to be empirically investigated.

2.2   Assessment Measures of Affect Expressed in Play Therapy

Even though there are many developmental measures to assess children's pretend play skills, there is relatively little evidence-based support for assessment measures that have been developed specifically to assess affective process and change in child play therapy. In particular, self-reported emotions are none too reliable, as they can be influenced by external factors [56].

Russ and Niec [54], in a review of play therapy assessment measures, talk about only three measures, which are Play Therapy Observation Instrument (PTOI) [28], the Trauma Play Scale [24] and the Children's Play Therapy Instrument (CPTI) [31], respectively. These are specifically designed to study children's expression of affect in therapy among other therapeutic indices. PTOI includes an Emotional Discomfort scale to rate child's comments about worries and troublesome events, inappropriate aggression toward the therapist, conflicted play, the quality and intensity of the child's affect (i.e., mood), and play disruption. The Trauma Play Scale allows for the coding of negative affect or lack of joy during play. CPTI has a more extensive affective component assessing affect regulation strategies as well as the types of affect expressed in play over the course of treatment. With all these measures, the sessions have to be recorded, transcribed and rated by trained judges on affective components.

2.3   Automatic Text Analyses of Affect from Text in Psychotherapy Research

A primary focus of the use of natural language processing (NLP) methods in psychotherapy has been to evaluate complex relational/emotional processes using the words from treatment sessions. Much of this work has involved the use of computerized dictionaries that place specific words in psychologically meaningful categories. For example, Anderson and colleagues found that when the patient used more emotion words, therapists obtained better outcomes when minimizing responses with cognitively geared verbs (e.g., "think," "believe," "know") [3]. Mergenthaler focused on the emotional tone (density of emotional words) and level of abstraction (the amount of abstract nouns) within patients' language and found that successful outcome in psychodynamic therapy is associated with

increased use of emotion and abstraction in language, which shows that the patients have emotional access to conflictual themes and can reflect upon them [38, 39].

Bucci's Referential Process theory is a similar, but more comprehensive psychological construct that "concerns the degree to which speakers (or writers) are able to access nonverbal, including emotional experience, in their own minds and to express this verbally in a form that is likely to evoke a corresponding experience in the listener" [9]). The affective connection between the language used and the underlying emotions has been consistently correlated with clinical ratings of psychoanalytic session effectiveness [10]. Pennebaker did not specifically investigate psychotherapy transcripts; however analyzed the writing features most strongly associated with enhanced psychological and physiological health found that people whose stories contained a high rate of what he called emotional processing words (e.g., "sad," "hurt," "guilt," "joy," "peace"), insight words (e.g., "realize," "understood," "thought," "know") and causal words (e.g., "because," "reason," "why") showed the greatest benefit from expressive writing exercises [45]. Even though there is substantial research in the application of NLP methods to specifically assess affective processes in adult treatment, to the best of our knowledge, no research has been carried out to adapt these measures to psychodynamic play therapy and there are no such resources in Turkish.

2.4   Text analysis for sentiment and affect detection

In multimedia computing, sentiment analysis and opinion mining refer to the categorization of a given text into positive, negative, or neutral classes, which makes it a relatively restricted and practicable NLP problem. On the other hand, detecting affect from text is a more challenging task, as it requires a profound understanding of both semantics and syntax of a language, as well as representing affect with the appropriate emotion categories or dimensions.

There exist several approaches to extract sentiment and opinion from textual multimedia content such as blogs, tweets, movie reviews and customer reviews. Basic methods include keyword spotting, lexical affinity, statistical NLP, learning based methods and commonsense-based approaches [44, 13]. Similarly, methods for affective content analysis from text generally blend these approaches with rule-based systems. An example is the Affect Analysis Model, which analyzes affect specifically in informal online communication media [43]. This approach has five main steps; symbolic cue analysis, syntactic structure analysis, word-level, phrase-level and sentence-level analysis, respectively.

The majority of research on affect analysis from text relies on lexicon-based approaches, in which a set of keywords and associated affect categories are used to generate features for affect prediction models. One of the comprehensive lexical resources in this area is the Affective Norms for English Words (ANEW) corpus [6], which includes a set of normative emotional ratings for 1,034 commonly used English words. This tool represents a set of verbal materials that have been rated in terms of pleasure, arousal, and dominance to support emotion

studies. Similarly, WordNet-Affect is a well known linguistic resource for extracting emotions from text [62]. The starting point of WordNet-Affect is to build a hierarchy of affective domain labels by labeling synsets (a set of one or more synonyms) that express affective concepts based on WordNet Domains [34].

A powerful system for text analysis is Linguistic Inquiry and Word Count (LIWC), which has a comprehensive affective dictionary to analyze text based on grammatical, psychological, and content word categorization. This dictionary allows to measure 74 different linguistic dimensions with more than 2,200 words and word stems. Affect sensing methods that are based on LIWC calculate word counts in the input text depending on these linguistic dimensions [30, 27, 46].

In addition to these lexicon-based approaches, several alternative methods have been studied in textual affect analysis. For example, Liu et al. first proposed the Commonsense-based approach by using three real-world commonsense databases [33]. Brooks et al. [8] presented an automated affect classification system in chat logs exploiting NLP and machine learning techniques. Their system segments the chat data and makes use of an improved bag-of-words model to classify text into 13 affect categories. The basic drawback of machine learning approaches is that they usually lack linguistic analysis by mainly focusing on statistical and syntactical features.

Recent approaches to text-based sentiment analysis rely on co-occurrence statistics, and in a multimedia context, typically combine image analysis with text [65]. To derive fixed length descriptors from variable length text fragments, the unsupervised Paragraph Vector approach proposed by Le and Mikolov is frequently used [32]. N-gram based generative approaches have shown some promise [40]. An example work for rule-based systems is Vader, which is tailored for social media text [29]. A recent review encompassing many application domains is given in [42].

3   The Proposed Text Analysis System

The automated affect analysis tool that is used in this work is designed to analyze affect and sentiment in Turkish online communication texts across domains [4]. Because of the lack of comprehensive Turkish corpora for affect analysis, we use an affect lexicon which is adapted from English to Turkish. English lemmas were gathered from the study of Warriner et al. [64], which evaluated 13,915 English lemmas in a nine point scale (1-9) by 1,827 participants through Mechanical Turk. For each item, mean and standard deviation values for valence, arousal, and dominance scores are given. Our text analysis model linearly transforms these affect scores to a five point scale [1-5]. Mapping to this range makes the scores given by the system directly comparable to the CPTI scores. The affective lexicon was expanded with synonym sets (synsets) from a standard Turkish dictionary (by TDK, Turkish Language Organization). As a result, a comprehensive affective lexicon for Turkish is developed, which includes valence, arousal, and dominance scores for 15,222 different words and phrases. We note here that

the translation process naturally introduces errors, and ignores cultural aspects entirely. Nonetheless, this approach produces a useful resource with little cost.

To deal with written communication, the model uses additional resources, including 120 emoticons, 98 abbreviations, 50 interjections, and 71 modifiers (emotion intensifiers and diminishers). The affect analysis model of the tool is illustrated in Figure 1.

In order to calculate sentence-level affect scores, the system first calculates the affective values of small units in the sentence, such as words and phrases, by tokenizing the sentence into trigrams, bigrams and unigrams. Next, the system checks the modifier list. If there is any modifier connected to a verb or a noun as phrasal, the score of the word is updated based on the polarity of the sentence and on the particular coefficient of the modifier. Then, the system handles negation and some morphological alternations and updates the valence and arousal scores accordingly.

The system exploits some linguistic rules when calculating the overall sentence score, based on simplifying assumptions. For example, considering the transitive verbs in Turkish, for NN+VB structures, such as "hayatını kaybetti" (he lost his life), only the affective score of the verb is taken and then the noun is neutralized. Similarly, if there is a NN+ADJ structure such as "kafam karışık" (I'm confused), the noun is neutralized and only the adjective is taken into consideration. The overall sentence score is computed by summing the scores of these units. Only words with affective load are considered in the summation.
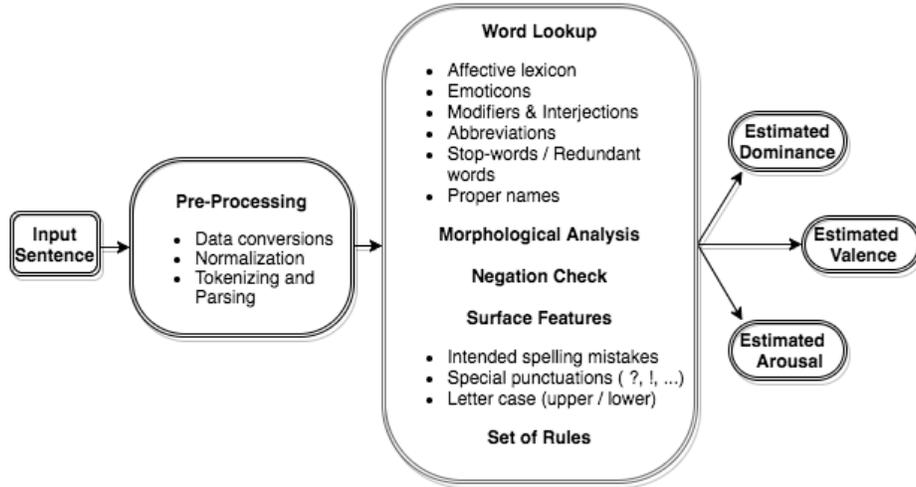

Fig. 1. Overview of the affect analysis system.

3.1   Adaptations for psychotherapy

The initial design of this system targeted general online communications [4]. As a part of this work, we adapted the system to the psychotherapy domain by updating the affect dictionary. During the translation of the dictionary, the primary meanings were used for each word, but synonyms were also stored as alternatives. We checked approximately 1,500 words manually and selected the word with the most appropriate sense in the psychotherapy domain and discarded the others. Another feature that we added to the system is the detection of frequent stop-words and redundant words in play therapy. For example, words such as "anne" (mother), "baba" (father), "oyun" (play) have high valence scores in our dictionary, however, these words are mostly present with a neutral tone during the play sessions. Therefore, we neutralized the affect scores of these words when calculating the overall affect score. Stop-word and redundant word lists include more than 500 words that we have treated as neutral words. We suggest that to adapt the system for a different domain, expert knowledge should be integrated at this level. The resources and code developed for this work is made available.

We next describe the experimental setup. We evaluate our approach on data collected during psychotherapy sessions, and contrast our findings with those of the expert psychotherapists.

4   Experimental Setup

We describe the experimental setup somewhat extensively here; the reader may skip to Section 4.3 for the details of data analysis and results.

4.1   Data

Patients The source of data used for this study comes from the Istanbul Bilgi University Psychotherapy Research Laboratory, which provides low-cost outpatient psychodynamic psychotherapy and professional training at master's level for students in the Clinical Psychology Program. Referrals were made by parents themselves or by mental health, medical, and child welfare professionals. The parents and the children were interviewed in order to determine whether the patients fit the study protocol inclusion criteria: ages between 4-10 years old; average intelligence; motivation for treatment; no psychotic symptoms; no significant developmental delays; no significant risk of suicide attempts; no drug abuse. The patients and their parents were extensively informed before commencing therapy and consented to video recordings and data collection at all times. The parents provided written informed consent and the children provided oral assent concerning use of their data for research purposes.

From September 2014 to September 2015, a group of 26 consecutively admitted patients who met inclusion criteria and consented to research were included in the study. 20 patients (76 %) completed the treatment. The demographics of the children are presented in Table 1. Eighty to ninety percent of the children

come from low to middle socioeconomic status (SES) families and approximately 10% of the parents are divorced or widowed for both samples. Referral Problems manifested primarily as anger management issues and behavioral problems, such as disobedience and not taking limits, followed by academic issues such as inattention in class and low grades and finally relational problems such as difficulties in family relationships or socialization with friends. At intake, 5 patients had DSM IV Attention Deficit and Hyperactivity Disorder, 3 patients had a Mood Disorder, 3 had Separation Anxiety Disorder, 2 patients had Encopresis.

Table 1. Subject Characteristics at Intake

Subject Characteristics at Intake		(N=20)
Sex	Male	9
	Female	11
Age		
	4-6 yrs.	8
	7-9 yrs.	12
SES		
	Low	4
	Low - Middle	7
	Middle	5
	Middle - High	4
Referral Problem		
	Anxiety Issues	2
	Behavioral/Anger Problems	10
	Academic Problem	7
	Relational Problems	1

Therapists A total of 12 therapists (all clinical psychology master's level graduate students) treated the 20 patients, with each therapist generally working with one to two patients. The therapists were all females with ages ranging from 23 to 27 years. Each therapist was extensively educated in the theoretical background of psychodynamic play therapy and its various applications one year prior to the study. All therapists had the same experience level (1-2 years of psychotherapy training) and were supervised by experienced clinicians. In this way, the confounding variables rooted in differences in the educational background, experience, and supervision process were partially controlled.

Treatment The treatment was psychodynamic play therapy. The treatment was not manualized and the only restrictions placed were regularity and length (once weekly treatment of 50 minutes for one year). Patients on average received

40 sessions. Even though there is no unitary model of therapeutic action in psychodynamic play therapy [25], the core principles and techniques employed can be summarized as follows: Central to this approach is the establishment of what is called a "setting". The psychotherapist sees the child at regular times, in the same play room with a standard set of play toys. This consistency provides a safe context that allows the child to play out difficult and disturbing emotional experiences that would be hard to express in the outside world. The exploration of the child's issues takes place in a largely child-led process way and the therapist encourages the child to express and reflect on his perceptions, feelings and thoughts in play. This is done by listening actively and inviting the child to continue his communications and asking questions about the play setting, temporal ordering, and the details of the characters, their thoughts, feelings and behaviors. The therapist also labels the repetitive themes, conflicts and feelings in play with the aim of helping the child to synthesize his experience. Interpretations aim to help the child see links between conflicting needs and emotions about self and others that find reflection in play behaviors and in the therapeutic relationship with the purpose of bringing to consciousness attitudes, assumptions and beliefs of which the child is unaware.

Session Selection For correlational analyses, the longest play segments of the first two sessions of psychotherapy were used. A total of 40 sessions and 40 play segments constituted the data points for the analysis. To run Multi-level Modeling and Trend Analyses, a total of six sessions were selected from each case. To represent different therapy phases, the sessions were divided into early, middle, and late phase by dividing the total number of sessions for each case by three. Two consecutive sessions were selected from early therapy, two from late therapy, and two from the middle. Each session included 1 to 10 play segments (see Section 4.2), with a mean of 2.3. Up to four play segments were selected from each session in order to achieve a balance among participants, since the number of play segments per session varied. This sampling resulted in 120 sessions and 289 play segments for 20 children.

4.2   Measures

Background information Demographic information such as socioeconomic status and marital status were obtained using a standard intake information form and from information obtained in the initial interview.

Outcome Measures The Child Behavior Checklist (CBCL) is a widely used method of identifying problematic behaviors in children [1]. For children ages 4 through 18, a parent or a primary caregiver reports on the child's academic performance, social relationships, and indicates how true a series of 112 problem behavior items are for the child on a 3 point scale (0 = not true, 1 = somewhat or sometimes true, and 2 = very true or often true). The following eight syndromes are scored from the CBCL, Anxious/Depressed, Withdrawn/

Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Rule Breaking Behavior, Aggressive Behavior. Anxious/Depressed, Withdrawn/Depressed, and Somatic Complaints syndromes comprise an Internalizing group, and the Rule Breaking Behavior and Aggressive Behavior syndromes comprise an Externalizing group, and Total Problems is the sum of scores on all problem items. The cut-off points for borderline and clinical designation are based on t–scores formed on a clinical population. Back translation, bilingual retest method, and pretest studies were used for the translation of the CBCL [22]. The test–retest reliability of the Turkish form was .84 for the Total Problems, and the internal consistency was adequate (Cronbach's alpha = .88; [21, 22]).

Assessment of Affect in Play Activity Children's Play Therapy Instrument (CPTI) is a psychodynamically-informed measure of in-session play activity [31]. The selected scales of the instrument for the purposes of the study involve Segmentation and Affects Expressed in Play (for further definition of play activity categories, see [15]). The CPTI rates children's behavior in a therapeutic setting at different levels. The first level involves a "Segmentation of the child's activity" (non-play, pre-play, play and interruption). Going forward, only play segments are rated. The Affective Component looks at the types emotions brought by the child to his play. Eight types of emotions are rated using a 5-point Likert scale: 5 = Most Characteristic; 4 = Considerable Evidence; 3 = Moderate Evidence; 2 = Minimal Evidence; 1 = No Evidence. For the purposes of the study, only Anger, Anxiety and Sadness were coded. Two masters level clinical psychology students, who received 20 hours of training on the CPTI by the first author and rated 10 training sessions (24 play segments) prior to the study, rated the sessions. They were independent assessors who were not associated with the treating clinicians or the cases, and blind to the purposes of the study. In order to identify the agreement level between judges for subscale ratings, Intra-class Correlation Coefficients (ICC) were computed. Cronbach Alpha was .72 for Segmentation, and .81 for Affect Types, suggesting good reliability for all Scales of CPTI.

Valence and Arousal Categorical and dimensional modeling are two main approaches in representation of affect [12]. In dimensional modeling, the assumption is that emotions are related to each other and the affective state is investigated in a continuous multidimensional space, in generally two or three dimensions. There is still a lack of consensus on which dimensions are fundamental and which dimensions are a mixture of these basic dimensions. However, the popular Circumplex model of emotions [49], which defines "valence" and "arousal" as the principal axes, is frequently used. Valence describes the extent of pleasure (positive) and sadness (negative), and arousal (or activation) describes the extent of calmness and excitation [49, 55]. Valence and arousal are commonly considered as independent dimensions, however, real-world findings confirm that these two dimensions are correlated most of the time.

4.3   Method of automatic analysis

As a general rule, linguistic programs need to segment the transcript (typically in equally sized units) for comparison of the data while analyzing a text. As the proposed text analysis tool performs sentence level analysis, firstly we had to segment sessions into smaller units. The length of a scoring unit containing the minimum number of necessary words is determined by statistical procedures described before [36]. In psychotherapy research, an entry with minimum of 150 words is required by many linguistic programs such as the therapeutic Cycle Model and computer-assisted content analysis [37]. Therefore, for the grouping, we created 150-word chunks of sentences while paying attention to play segment borders. Each 150 word block was processed as a single sentence in our affect analysis system. Then, average scores of these 150 word blocks gave us the overall affect score of the corresponding therapy session.

5   Results

5.1   Descriptive Statistics

To examine the association between CBCL problems and affect expressed in play at the beginning of therapy, play affect scores as measured by CPTI Anger, Anxiety and Sadness scores, VA (Valence and Arousal) scores collected in the initial two sessions of psychotherapy were calculated. Each child's two longest play segment affect scores from the initial two sessions were computed, which gave mean affect scores for the initial phase of psychotherapy. The means and standard deviations for each of the major variables collected at the beginning of psychotherapy are listed in Table 2. The first two rows (Valence and Arousal) are obtained by the proposed automatic analysis approach, and the next three rows are CPTI annotations (Anger, Anxiety and Sadness).

Prior to testing correlations, the possible contribution of background and demographic variables to the studied variables was examined through preliminary analyses. Spearman correlations were conducted to assess the association of age and gender with the main study variables: CBCL Problems and all CPTI Items and VA. No significant differences were found according to these variables.

5.2   Preliminary results of affect analysis at the beginning of treatment

The relationship between the CBCL Problems and play affect scores as measured by CPTI Anger, Anxiety and Sadness scores and Valence and Arousal scores collected in the initial two sessions of psychotherapy were examined. Due to the low number of children included in the analysis, Spearman Correlations were used (see Table 2).

Results show that in the first two sessions, CPTI Anger scores were positively related to Externalizing Problems, Valence scores were negatively related to Internalizing and Externalizing Problems, and Arousal Scores were negatively

Table 2. Descriptive Statistics for Affect Variables and CBCL Problems (N =20)

	Affect Variables	
Variable	**M**	**SD**
Valence	3.48	0.38
Arousal	3.54	0.71
Anger	2.71	0.87
Anxiety	3.03	0.61
Sadness	1.4	0.42
	CBCL Problems (T Scores)	
Externalizing Problems	61.80	8.70
Internalizing Problems	59.95	11.14

related to Internalizing Problems on the CBCL. No significance was observed for CPTI Sadness and Anxiety scores (see Table 3).

Table 3. Spearman Correlations between the affect scores and CBCL Problems.

	CPTI Anxiety	CPTI Sadness	CPTI Anger	VA Valence	VA Arousal
CBCL					
Internalizing Prob.	-.343	.218	-.014	-.495*	-.644**
Externalizing Prob.	-.011	-.025	.496*	-.517*	.253

Note: Correlation is significant at the .05 level; ** Correlation is significant at the .01 level.

While we do not analyze the specific findings of the play therapy sessions in detail here, we note that the high correlations obtained by the proposed automatic tool and the manual CPTI coding are very promising. The results provide empirical support for two measures of affective assessment that can be used towards investigating affective processes in play in psychodynamic play therapy. Both CPTI and Valence-Arousal showed preliminary promise for systematic play observation.

5.3   Preliminary Analyses of Affect Expressed During Treatment

In order to assess affect expressed during the treatment, two sessions from the beginning, middle and end of therapy were used. As such the data consisted of 6 sessions from 20 children resulting in 120 sessions and 289 play segments. We conducted Hierarchical Linear Modelling (HLM) [50] which is used to measure data that has more than one level. Using Hierarchical Linear Growth Curve Modeling, affective change over time was modeled. This model takes into account the hierarchical structure of the data i.e., different measurements in time (level

1) are nested within subjects (level 2). Using maximum likelihood, multilevel analysis allows for missing data [60]. Effect-sizes were calculated using R2.

To see the variability of mean valence, arousal and anger scores, first null models were run for each. Results showed that Valence (β= 3.39, t(19) = 124.36, p <0.001), Arousal (β = 3.54, t(19) = 41.64, p <0.001) and Anger (β = 2.59, t(19) = 18.09, p <0.001) significantly varied across participants.

Results also revealed that left over variance was significant for Arousal (Var(u_0) = 0.08, χ^2 (19) = 49.49, p <0.001), Anger (Var(u_0) = 0.27, χ^2 (19)= 54.36, p <0.001), but not for Valence (Var(u_0)= 0.00, χ^2(19)=15.34, p >0.05). We also calculated how much of the variance is explained by level 2 variables (Externalizing and Internalizing problems) in predicting Valence, Arousal and CPTI Anger. To calculate this we used the following formula:

$$Explained\ variance = \frac{u_0\ (unconditional) - u_0\ (conditional)}{u_0\ (unconditional)} \tag{1}$$

Because HLM does not give a direct R2 value, the variance explained with this formula can be used as pseudo R2 [2].

We found that for Arousal 16% and for Anger 19% of the variance is explained by Externalizing and Internalizing Problems. For Valence, we could not obtain a value because left over variance was not significant at null model as stated above. Together, these results indicated that further analysis using Hierarchical Linear Modeling, was suited.

Growth Curve Analyses To investigate the change in Internalizing and Externalizing children's Valence, Arousal and Anger scores across sessions, time and time squared variables were entered into the model at level 1 to see the linear and quadratic growth rates of variables. For Internalizing Problems, results revealed a significant linear increase (β = 0.05, t(161) = 2.98, p <0.05) in Valence as well as Arousal (β = 0.06, t(161) = 2.17, p <0.05) scores. Effect sizes for each trend was small (R2 = 0.01). No significance was observed for linear (β = -0.02, t(161) = 0.18, p >0.05) and quadratic effect for CPTI Anger (β = -0.03, t(161) = 0.58, p >0.05). Growth rates of Valence, Arousal and CPTI Anger with Externalizing Problems were not significant.

5.4   Ablation study

We assess the impact of different parameters on the accuracy of our affect prediction system. To achieve that, we setup a sentence-level annotation with 4 different play therapy sessions that includes approximately 500 sentences in total. For each sentence, a human annotator assigned a Valence and an Arousal score by using a 5-point Likert scale. After the automated affect analysis, we compared the model prediction scores with the ground truth scores that we obtained from the annotation. Model scores are also scaled continuously between 1 and 5. In order to calculate the accuracy, we mapped the Valence scores to positive (>3) and negative (<3) classes to carry out the corresponding classification of the affect.

The first experiment we conducted tested the benefit of using domain adaptation on the text analysis system. As can be seen from Table 4, with the updated dictionary and redundant word elimination, we observed a higher correlation and reduced mean square error in both Valence and Arousal dimensions.

Table 4. The effect of adapted dictionary for psychotherapy domain

	Valence		Arousal	
Measure	**Adapted Dict.**	**Generic Dict.**	**Adapted Dict.**	**Generic Dict.**
Correlation (P <0.01)	0.58	0.32	0.33	0.23
Mean Square Error	0.24	0.39	0.51	0.58

Contribution of the different parts of the system to the performance is given in Table 5. Our results show that the system gives the best accuracy (83%) when all features are employed with the adapted dictionary for psychotherapy domain. We see that eliminating the domain-specific redundant words improves the system performance by 2%.

Table 5. The accuracy of the model for binary Valence classification

	Accuracy (%)
All features	**83.5**
All features with generic dictionary	74.5
All features without redundant words	81.1
All features without negation	75.1
All features without modifiers	79.8

6   Conclusions

There is relatively little empirical investigation of the measurement of affect expressed in play and how it relates to psychopathology during the treatment process of children in psychodynamic play therapy. We propose in this paper an automatic rule-based text analysis tool that can quantify Valence and Arousal for longitudinal transcriptions of therapy sessions. We obtain good agreement with a standard measure used by psychotherapists. Result of the study support the relationships between affect expressed in play and behavioral problems as well as the importance of play in the modulation of negative feelings. The findings were consistent with our prediction which indicated Internalizing and Externalizing Problems negatively associated with Valence at the beginning of

treatment. These findings parallel previous results from the literature that suggest a relationship between negative affect in play and maladaptive behavior. Our findings also indicated, in line with previous literature, that children with Internalizing Problems present with a constricted range of negative affect and can use psychodynamic play therapy towards the modulation of negative affect in play. They are able to express more intense and positive emotions over the course of treatment as shown in the increase in Arousal scores. These findings provide preliminary empirical support for two measures of affective assessment that can be used towards investigating affective processes in play in psychodynamic play therapy.

One of the main limitations of the study is that none of the existing text-based sentiment analysis approaches could be directly employed for comparative assessment, as few approaches are proposed for Turkish (see [4] and references therein). It is obvious that improvements in the automatic affect analysis pipeline will translate to more reliable assessment of the play therapy sessions. In particular, a comprehensive affective lexicon prepared for Turkish language would be useful. The work by Dehkharghani et al. towards preparing such a resource is a good step forward [18], but currently it is in a preliminary stage, and the translated (but more extensive) dictionary we have used produces more accurate results [4].

Our work also indicates that it is possible to adapt sentiment analysis resources developed for one language (i.e. English, in this case) for a system designed for processing another language. N-gram and co-occurrence based approaches do not have this flexibility, and need to be trained directly with resources of the language they are meant to process. Subsequently, the proposed approach presents a possibility of supporting and complementing these methods.

7   Acknowledgments

References

1. T. M. Achenbach. *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Department of Psychiatry, University of Vermont Burlington, VT, 1991.
2. D. Anderson. Hierarchical linear modeling (hlm): An introduction to key concepts within cross-sectional and growth modelling frameworks. *Oregon: Behavioral Research & Teaching*, 2012.
3. T. Anderson, E. Bein, B. Pinnell, and H. Strupp. Linguistic analysis of affective speech in psychotherapy: A case grammar approach. *Psychotherapy Research*, 9(1):88–99, 1999.
4. E. Aydın Oktay, K. Balcı, and A. A. Salah. Automatic assessment of dimensional affective content in Turkish multi-party chat messages. In *Proceedings of the International Workshop on Emotion Representations and Modelling for Companion Technologies*, pages 19–24. ACM, 2015.

5. L. A. Barnett and B. Storm. Play, pleasure, and pain: The reduction of anxiety through play. *Leisure Sciences*, 4(2):161–175, 1981.

6. M. M. Bradley and P. J. Lang. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical report, C-1, The Center for Research in Psychophysiology, Univ. of Florida, 1999.

7. S. C. Bratton, D. Ray, T. Rhine, and L. Jones. The efficacy of play therapy with children: A meta-analytic review of treatment outcomes. *Professional Psychology: Research and Practice*, 36(4):376, 2005.

8. M. Brooks, K. Kuksenok, M. K. Torkildson, D. Perry, J. J. Robinson, T. J. Scott, O. Anicello, A. Zukowski, P. Harris, and C. R. Aragon. Statistical affect detection in collaborative chat. In *Proc. CSCW*, pages 317–328. ACM, 2013.

9. W. Bucci and B. Maskit. Beneath the surface of the therapeutic interaction: The psychoanalytic method in modern dress. *Journal of the American Psychoanalytic Association*, 55(4):1355–1397, 2007.

10. W. Bucci, B. Maskit, and S. Murphy. Connecting emotions and words: the referential process. *Phenomenology and the Cognitive Sciences*, pages 1–25, 2015.

11. J. L. Butcher and L. N. Niec. Disruptive behaviors and creativity in childhood: The importance of affect regulation. *Creativity Research Journal*, 17(2-3):181–193, 2005.

12. R. A. Calvo and S. Mac Kim. Emotions in text: dimensional and categorical models. *Computational Intelligence*, 29(3):527–543, 2013.

13. E. Cambria, B. Schuller, Y. Xia, and C. Havasi. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2):15–21, 2013.

14. R. J. Casey. Emotional competence in children with externalizing and internalizing disorders. *Emotional development in atypical children*, pages 161–183, 1996.

15. S. Chazan. *Profiles of play: Assessing and observing structure and process in play therapy.* Jessica Kingsley Publishers, 2002.

16. M. Chethik. *Techniques of child therapy: Psychodynamic strategies.* Guilford Press, 2003.

17. L. D'Angelo. Child's play: The relationship between the use of play and adjustment styles. *Unpublished Doctoral Dissertation*, 1995.

18. R. Dehkharghani, Y. Saygin, B. Yanikoglu, and K. Oflazer. SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation*, pages 1–19, 2015.

19. J. Dunn and C. Hughes. I got some swords and you're dead!: Violent fantasy, antisocial behavior, friendship, and moral sensibility in young children. *Child Development*, 72(2):491–505, 2001.

20. N. Eisenberg, A. Cumberland, T. L. Spinrad, R. A. Fabes, S. A. Shepard, M. Reiser, B. C. Murphy, S. H. Losoya, and I. K. Guthrie. The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. *Child Development*, 72(4):1112–1134, 2001.

21. N. Erol, B. Arslan, and M. Akçakın. The adaptation and standardization of the child behavior checklist among 6–18 year-old Turkish children. *Eunethdis: European Approaches to Hyperkinetic Disorder*, pages 97–113, 1995.

22. N. Erol and Z. T. Şimşek. Mental health of Turkish children: Behavioral and emotional problems reported by parents, teachers, and adolescents. *International Perspectives on Child and Adolescent Mental Health*, 1:223–247, 2000.

23. G. G. Fein. Pretend play: Creativity and consciousness. In P. Gorlitz and J. Wohlwill, editors, *Curiosity, imagination, and play*, pages 281–304. Hillsdale, NJ: Lawrence Erlbaum Associates, 1987.

24. J. H. Findling, S. C. Bratton, and R. K. Henson. Development of the trauma play scale: An observation-based assessment of the impact of trauma on play therapy behaviors of young children. *International Journal of Play Therapy*, 15(1):7, 2006.

25. P. Fonagy and M. Target. The place of psychodynamic theory in developmental psychopathology. *Development and Psychopathology*, 12(03):407–425, 2000.

26. T. J. Gaensbauer and C. H. Siegel. Therapeutic approaches to posttraumatic stress disorder in infants and toddlers. *Infant Mental Health Journal*, 16(4):292–305, 1995.

27. J. T. Hancock, C. Landrigan, and C. Silver. Expressing emotion in text-based communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 929–932. ACM, 2007.

28. P. A. Howe and L. E. Silvern. Behavioral observation of children during play therapy: Preliminary development of a research instrument. *Journal of Personality Assessment*, 45(2):168–182, 1981.

29. C. J. Hutto and E. Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

30. J. H. Kahn, R. M. Tobin, A. E. Massey, and J. A. Anderson. Measuring emotional expression with the linguistic inquiry and word count. *The American journal of psychology*, pages 263–286, 2007.

31. P. Kernberg, S. Chazan, and L. Normandin. The children's play therapy instrument (CPTI). description, development, and reliability studies. *The Journal of Psychotherapy Practice and Research*, 7(3):196–207, 1997.

32. Q. V. Le and T. Mikolov. Distributed representations of sentences and documents. In *ICML*, volume 14, pages 1188–1196, 2014.

33. H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *Proc. IUI*, pages 125–132. ACM, 2003.

34. B. Magnini and G. Cavaglia. Integrating subject field codes into WordNet. In *LREC*, 2000.

35. A. K. Marcelo and T. M. Yates. Prospective relations among preschoolers' play, coping, and adjustment as moderated by stressful events. *Journal of Applied Developmental Psychology*, 35(3):223–233, 2014.

36. E. Mergenthaler. Textbank systems. *Computer science applied in the field of psychoanalysis*, 1985.

37. E. Mergenthaler. Computer-assisted content analysis. *Nachtrichten Spezial*, pages 3–32, 1996.

38. E. Mergenthaler. Emotion–abstraction patterns in verbatim protocols: A new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology*, 64(6):1306–1315, 1996.

39. E. Mergenthaler. Cycles of emotion-abstraction patterns: A way of practice oriented process research? *The British Psychological Society-Psychotherapy Section Newsletter*, 24:16–29, 1998.

40. G. Mesnil, T. Mikolov, M. Ranzato, and Y. Bengio. Ensemble of generative and discriminative techniques for sentiment analysis of movie reviews. *ICLR*, 2014.

41. M. E. Milos and S. Reiss. Effects of three play conditions on separation anxiety in young children. *Journal of Consulting and Clinical Psychology*, 50(3):389, 1982.

42. S. M. Mohammad. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. *Emotion Measurement*, 2016.

43. A. Neviarouskaya, H. Prendinger, and M. Ishizuka. Affect analysis model: novel rule-based approach to affect sensing from text. *Natural Language Engineering*, 17(01):95–135, 2011.

44. B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

45. J. W. Pennebaker and C. K. Chung. Expressive writing, emotional upheavals, and health. *Foundations of health psychology*, pages 263–284, 2007.

46. J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54(1):547–577, 2003.

47. R. D. Phillips. Whistling in the dark? a review of play therapy research. *Psychotherapy: Theory, Research, Practice, Training*, 22(4):752, 1985.

48. R. Plutchik. *Emotions in the practice of psychotherapy: Clinical implications of affect theories.* American Psychological Association, 2000.

49. J. Posner, J. A. Russell, and B. S. Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(3):715–734, 2005.

50. S. Raudenbush, A. Bryk, M. Seltzer, and R. Congden. *An introduction to HLM: Computer program and user's guide.* Chicago: Department of Education, University of Chicago, 1986.

51. S. Russ and M. Cooperberg. Play as a predictor of creativity, coping and depression in adolescence. *Manuscript submitted for publication*, 2002.

52. S. W. Russ. *Affect and creativity: The role of affect and play in the creative process.* Psychology Press, 1993.

53. S. W. Russ. *Play in child development and psychotherapy: Toward empirically supported practice.* Routledge, 2003.

54. S. W. Russ and L. N. Niec. *Play in clinical practice: Evidence-based approaches.* Guilford Press, 2011.

55. J. A. Russell. Culture and the categorization of emotions. *Psychological bulletin*, 110(3):426, 1991.

56. S. Schachter and J. Singer. Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69(5):379–399, 1962.

57. T. J. L. Scott, E. J. Short, L. T. Singer, S. W. Russ, and S. Minnes. Psychometric properties of the dominic interactive assessment a computerized self-report for children. *Assessment*, 13(1):16–26, 2006.

58. S. R. Shirk and R. L. Russell. *Change processes in child psychotherapy: Revitalizing treatment and research.* Guilford Press, 1996.

59. D. G. Singer and J. L. Singer. *The house of make-believe: Children's play and the developing imagination.* Cambridge, MA: Harvard University Press, 1990.

60. J. D. Singer and J. B. Willett. *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford university press, 2003.

61. J. L. Singer. Imaginative play in childhood: Precursor of subjunctive thoughts, daydreaming, and adult pretending games. *The future of play theory*, pages 187–219, 1995.

62. C. Strapparava and A. Valitutti. WordNet Affect: an affective extension of WordNet. In *LREC*, volume 4, pages 1083–1086, 2004.

63. K. Von Klitzing, K. Kelsay, R. N. Emde, J. Robinson, and S. Schmitz. Gender-specific characteristics of 5-year-olds' play narratives and associations with behavior ratings. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39(8):1017–1023, 2000.

64. A. B. Warriner, V. Kuperman, and M. Brysbaert. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods*, 45(4):1191–1207, 2013.

65. Q. You, J. Luo, H. Jin, and J. Yang. Joint visual-textual sentiment analysis with deep neural networks. In *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, pages 1071–1074, New York, NY, USA, 2015. ACM.