

Fiabilidad de un instrumento de evaluación emocional a través de la concordancia inter-observador para el aprendizaje basado en juegos

Carina Soledad González González
Departamento de Ingeniería
Informática y de Sistemas
Universidad de La Laguna
San Cristóbal de La Laguna, España
cjgonza@ull.edu.es

Vicente Navarro Adelantado
Departamento de Didácticas
Especiales
Universidad de La Laguna
San Cristóbal de La Laguna, España
vnavarro@ull.edu.es

Mariana Cairós González
Facultad de Psicología
Universidad de La Laguna
San Cristóbal de La Laguna, España
mariana.cairos@gmail.com

Abstract — *En este trabajo se presentan los resultados del cálculo de fiabilidad de un instrumento de medición subjetiva de emociones en niños y niñas, al cual hemos denominado EMODIANA. La EMODIANA tiene tres dimensiones de análisis: la identificación de una emoción con una expresión facial subjetiva, la identificación de la intensidad de la emoción y la justificación o explicación de la emoción identificada. Este trabajo se centra en el análisis cualitativo de los resultados de las justificaciones realizadas por los menores. Para ello, se estudió el grado de acuerdo entre tres jueces para cada ítem de valoración subjetiva de la muestra utilizando el coeficiente de Kappa de Fleiss para tres observadores. Nuestro coeficiente Kappa alcanza un valor $k=0,903$. La alta fiabilidad del sistema de categorías acredita que el instrumento aplicado pueda ser utilizado para la investigación de la evaluación emocional en niños y niñas en programas de intervención educativa basadas en juegos.*

Keywords — *fiabilidad entre codificadores, Kappa de Fleiss, sistema de categorías, instrumento de evaluación emocional, Evaluación de emociones en niños y niñas, Aprendizaje basado en juegos.*

I. INTRODUCCIÓN

La evaluación emocional intenta recopilar y medir información sobre aspectos cualitativos y cuantitativos de la experiencia de un usuario. El instrumento de medición dependerá en gran medida de lo que se busca obtener, pudiendo distinguir entre instrumentos no verbales (objetivos) y verbales (subjetivos). Actualmente existen distintas técnicas de medición emocional, tales como, técnicas que permiten analizar las expresiones faciales, técnicas que miden reacciones fisiológicas (por ejemplo, latido del corazón, sudor, dilatación de la pupila o las ondas cerebrales) o las técnicas de medición subjetiva de sentimientos a través de cuestionarios, entrevistas y auto-informes (Ekman & Friesen, 1975; Kaiser & Wehrle, 2001; Schneider & Josephs, 1991; Picard & Daily, 2001; Bradley & Lang, 1994). Es en este último tipo de técnicas en donde situamos el instrumento de evaluación emocional que presentamos en este trabajo.

Existen distintos instrumentos que realizan este tipo de evaluaciones, entre los que encontramos SAM (Self-Assessment Manikin) (Bradley & Lang, 1994) o Premo© (Product Emotion

Measurement Instrument) (Desmet, 2005). Este último se basa a su vez en SAM, pero está especialmente diseñado para medir la experiencia emocional de los usuarios con productos. Aunque este método ha sido utilizado con éxito en niños, la cantidad de emociones que tiene (14) es potencialmente problemático para su uso como un método de auto-reporte durante una actividad (Lewis, Haviland-Jones & Barrett, 2008). Esto se debe a que la carga cognitiva que produce es significativa para los niños y niñas, ya que deben identificar, distinguir y seleccionar sus respuestas emocionales entre 14 emociones. Igualmente, existen dificultades de asociación entre el lenguaje utilizado por los

5

niños y niñas y su asociación con la representación gráfica de la emoción (Lewis, Haviland-Jones & Barrett, 2008). Por este motivo se necesita diseñar un nuevo instrumento que tenga en cuenta las dificultades cognitivas de los niños y niñas, y que se base de las fortalezas de los métodos anteriores.

En un trabajo anterior se presentó la elaboración y validación del instrumento EMODIANA (González, Cairós, Navarro, 2013). En el mismo, se analizó, por un lado, el diseño gráfico de un instrumento de evaluación emocional en referencia a la comprensión por parte de los niños y las niñas de las representaciones gráficas emocionales (expresiones faciales o caras) y por otra parte, la intensidad emocional asociada a cada emoción. Además, en el mismo trabajo se presentó el estudio sobre la coherencia entre el lenguaje utilizado al definir las emociones en Premo©, la literatura (Pons, 2003; Lewis, Haviland-Jones & Barrett, 2008; Padilla-Zea et al., 2012) y el lenguaje de los niños y niñas utilizado en la identificación de las emociones.

Por otra parte, la fiabilidad de un estudio de contenido se basa en aplicar un sistema de forma fiable. Esta fiabilidad pasa por hallar un grado de acuerdo significativo entre los observadores (o entre los codificadores) del contenido cualitativo. En este sentido, uno de los métodos más empleados para medir la concordancia entre codificadores u observadores es el estadístico de acuerdo porcentual, el cual refleja el número de acuerdo respecto al número total de codificaciones realizadas, como por ejemplo el coeficiente de fiabilidad de Holsti (1969)



(Torres & Perera, 2009). Otros investigadores manifiestan que el acuerdo entre codificadores puede ser una medida inadecuada ya que no se tiene en cuenta el azar entre investigadores entre la concordancia observada (Capozzoli, McSweeney & Sinha, 1999). Esto lo corrige el estadístico Kappa de Cohen (k) que se utiliza para dos codificadores, en n casos para m categorías nominales y exclusivas mutuamente.

Averiguar la fiabilidad del instrumento y la de los observadores es una cuestión única, por ello, en este trabajo presentaremos el análisis de la fiabilidad del instrumento EMODIANA utilizando la Kappa de Fleiss (Fleiss & Cohen, 1973), como prueba de concordancia adecuada para múltiples codificadores u observadores (Losada & Arnau, 2000; Cerda & Villaroel, 2008).

El instrumento EMODIANA ha sido empleado para la evaluación emocional de un programa de intervención educativo sobre hábitos saludables basado en juegos motores y videojuegos, a través del cual se han recogido los datos que fueron analizados para la realización de las pruebas de fiabilidad.

A continuación, se describe en primer lugar el instrumento EMODIANA, luego veremos los fundamentos teóricos de los coeficientes de fiabilidad utilizados. Posteriormente, presentaremos la metodología utilizada en el estudio (la muestra del estudio y el procedimiento seguido para analizar el instrumento el proceso de concordancia entre los codificadores), presentando los resultados obtenidos sobre la fiabilidad del sistema de categorías y finalmente, las conclusiones sobre los resultados.

II. INSTRUMENTO DE MEDICIÓN SUBJETIVA DE EMOCIONES EN NIÑOS Y NIÑAS

EMODIANA es un instrumento de evaluación emocional subjetiva para niños y niñas, con edades comprendidas entre 7 y 12 años. Este instrumento, basado en Premo©, fue creado para evaluar emociones en situaciones educativas con juegos motores y videojuegos. La validación inicial del instrumento se basó en el análisis del diseño gráfico y significado emocional atribuido a las diferentes expresiones gestuales presentadas a 168 niños y niñas de entre 7 y 12 años. Asimismo, se validó el lenguaje infantil utilizado para definir las emociones y el grado de discriminación emocional de los 168 menores participantes en el estudio.

Teniendo en cuenta los resultados obtenidos sobre las caras y el lenguaje infantil para definir las emociones que éstas representan, hemos definido las 10 categorías emocionales del nuevo instrumento de medición emocional para niños EMODIANA (Figura 1).

Asimismo, la EMODIADA, a diferencia de Premo©, utiliza de forma combinada las expresiones emocionales del personaje junto a la etiqueta correspondiente a la categoría emocional, de forma de asegurar la comprensión de la emoción elegida. De esta forma, la EMODIANA contribuye a minimizar los errores en la evaluación subjetiva de las emociones de los niños y niñas, debidos a la incompreensión de los elementos gráficos o del lenguaje utilizado. Además, esta herramienta incluye la

valoración de la intensidad emocional a través del uso de la diana (Figura 1), en donde el centro representa la intensidad más alta.

El procedimiento de utilización de la EMODIANA se divide en dos fases, según los momentos de la sesión en donde se aplicará el instrumento: fase pre-sesión y fase post-sesión. De esta forma, al entrar en la sesión, el evaluador/a presenta al niño la EMODIANA (en formato de cartel gráfico), le pregunta al niño/a cómo se siente y le pide que ubique un imán gráfico en la diana, diciéndole que en el centro de la diana es la mayor intensidad (mucho) y el exterior la menor intensidad (poco). Luego, le pide al niño/a que explique las causas o razones de por qué ha dicho que se siente de esa forma. Este procedimiento se registra en una hoja de registro emocional de la sesión, incluyendo valores cuantitativos (emoción e intensidad) y cualitativos (explicaciones dadas por los niños/as). Al terminar la sesión, el evaluador/a repetirá el procedimiento de registro, pidiéndole al niño/a que diga qué siente, en qué intensidad y por qué se siente así, y lo registrará en la hoja de registro de la sesión. El registro se realiza tal y como el contenido es aludido literalmente por los niños.

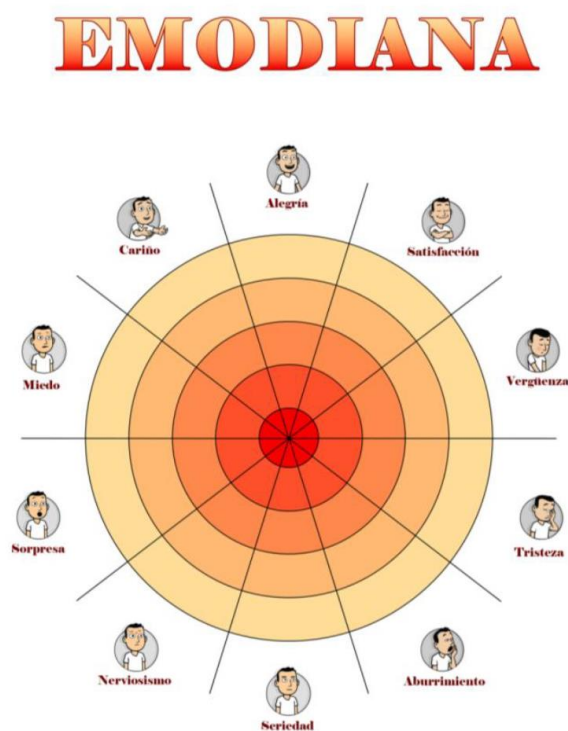


Fig. 1. Instrumento de evaluación emocional subjetiva para niños y niñas EMODIANA

La información recogida en las sesiones de intervención se analiza siguiendo procedimientos estadísticos para el análisis cuantitativo (medias, frecuencias, varianzas, etc.), y para el análisis cualitativo (codes y categorías, fiabilidad, y concordancia). En este trabajo se presenta el análisis de fiabilidad y concordancia de los codificadores.

III. FIABILIDAD DEL INSTRUMENTO

Lo esencial en cualquier tipo de instrumento de medida es que mida lo que efectivamente pretende medir, y que lo haga con

la máxima consistencia posible. El grado en que un instrumento mide lo que dice medir, viene indicado por su "validez" y su consistencia al hacerlo, por la "fiabilidad" (Torres & Perera, 2009). Así, validez y fiabilidad son las propiedades necesarias de todo instrumento de medida, en nuestro caso: la EMODIANA.

Por tanto, es posible construir un instrumento válido que nos cuantifique lo que deseamos medir, en nuestro caso las emociones de los niños/as antes y después de la sesión de intervención educativa con juegos y videojuegos. Para ello, debemos comprobar si la medida tomada es estable y si es válida para varios sujetos. Si nuestra medida no es constante, debemos preguntarnos si es debido a una falta de constancia intrínseca en nuestro instrumento, o a la falta de constancia intrínseca a la conducta humana. Por ello, nos preocupa estudiar estos problemas frente a la fiabilidad de nuestro instrumento.

La validez comprobada es aquella que ha sido verificada de alguna manera empírica o experimental. Puede ser empírica o teórica. Es empírica, cuando se ha comprobado de modo práctico, y teórica, cuando se ha comprobado con argumentos científicos que el test mide la propiedad que pretende medir. Para comprobar la validez empírica de un test es preciso definir un criterio externo de validez. Llamamos coeficiente de validez del test a la correlación entre el test y el criterio utilizado. Si los sujetos varían en el test como en el criterio, la correlación será elevada y el test tendrá un alto coeficiente de validez respecto a ese criterio. De esta manera, es evidente que el problema principal estriba en elegir un criterio lo más adecuado posible y en medirlo adecuadamente y con precisión.

La puntuación que un sujeto obtiene en un instrumento (por ejemplo, un cuestionario) está afectada por un error de medida, que se puede estimar a cierto nivel de confianza. Para ello, sirve el estudio de la fiabilidad. Si desconocemos la fiabilidad de un instrumento, sabemos que su medida no es precisa, pero no sabemos la magnitud de su imprecisión. Por el contrario, si la conocemos, podemos averiguar en qué medida el instrumento es impreciso y hasta qué punto podemos fiarnos de sus resultados. Con el objetivo de conocer el grado de validez del instrumento EMODIANA, este trabajo presenta el estudio de su fiabilidad basada en la concordancia entre los observadores.

Muestra

La muestra de nuestro estudio se conformó con un grupo escolar correspondiente del 3er al 6to curso de Educación Primaria de un colegio público de Tenerife. Este estudio se realizó en el marco de un proyecto de investigación sobre la intervención educativa con juegos motores y videojuegos dirigidos a la formación de hábitos saludables. Por ello, la muestra fue seleccionada al azar entre el grupo escolar que cumpliera las siguientes características: sobrepeso (IMC >95 y >97) y ordenador con conexión a Internet en sus casas. De esta forma, el tamaño de la muestra obtenida fue de 24 niños y niñas, 12 para el grupo control y 12 para el grupo experimental. La edad de los niños y niñas estudiados fue de 8 a 12 años. La EMODIANA fue utilizada solo con el grupo experimental. La aplicación de las pruebas se llevó a cabo en el colegio anteriormente mencionado durante el curso escolar, al final de una jornada escolar (de 15 a 17 horas).

Al llevar a cabo nuestra investigación, pusimos especial atención en controlar cualquier error que pudiera introducirse en ella a causa del procedimiento y su protocolo. De esta forma, la EMODIANA se utilizó siguiendo el procedimiento indicado en la sección 2, al principio y al final de la sesión. De un total de un programa de intervención educativa de juegos motores y videojuegos, compuesto de 16 sesiones, la EMODIANA se aplicó en las 6 sesiones finales. La aplicación de la prueba se hizo, como hemos dicho, de forma colectiva, en una sola sesión de dos horas aproximadamente.

Se mantuvo siempre constante el equipo de aplicadores, formando por dos personas, una psicóloga y una pedagoga, ambas familiarizadas con este tipo de trabajos y con conocimientos del instrumento y del procedimiento. Una de ellas, la misma en cada ocasión, fue la encargada de dar las consignas en todas las aplicaciones, actuando la otra persona como ayudante, estando, por lo tanto, encargado de supervisar, repartir y recoger los protocolos.

La prueba fue realizada en el espacio de la biblioteca del colegio, sin interferencias de otras personas, la cual era utilizada normalmente durante el curso y reunía las condiciones adecuadas para el trabajo que se iba a realizar. No se advirtió ninguna variable que influyese diferencialmente en alguna de las sesiones, manteniéndose constantes las condiciones generales de aplicación para todos los sujetos.

3.2. Procedimiento de análisis de datos

Luego de haber realizado las pruebas con la EMODIANA, pasamos al análisis de los datos literales obtenidos de las fichas de registro. Fueron cumplimentadas 12 fichas de datos, recogidas al inicio y final de cada sesión (6). Las puntuaciones de esas fichas fueron adecuadamente trasladadas a tablas para su posterior análisis estadístico en SPSS. Las emociones percibidas por los sujetos durante el total de las sesiones fueron clasificadas en positivas, neutras y negativas, observándose emociones altamente positivas (Figura 2) (Tabla 1).

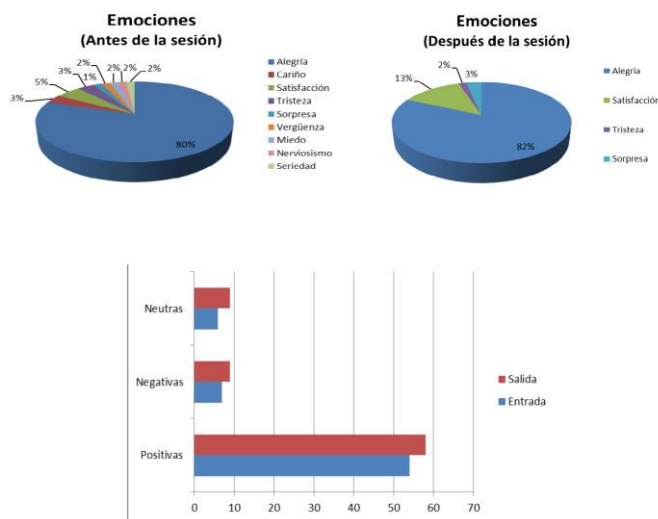


Fig. 2. Emociones declarativas categorizadas en positivas, negativas y neutras observadas antes y después de cada sesión.

Emoción	Llegada		Salida		%
	Tipo	Frecuencia	%	Frecuencia	
Alegría	Positiva	49	80,33	50	81,97
Cariño	Positiva	2	3,28	0	0,00
Satisfacción	Positiva	3	4,92	8	13,11
Tristeza	Negativa	2	3,28	1	1,64
Miedo	Negativa	1	1,64	0	0,00
Nerviosismo	Negativa	1	1,64	0	0,00
Seriedad	Neutra	1	1,64	0	0,00
Sorpresa	Neutra	1	1,64	2	3,28
Vergüenza	Neutra	1	1,64	0	0,00
Totales		61	100,00	61	100,00

Tabla 1. Emociones declarativas percibidas en las sesiones de juegos motores y videojuegos activos

La fiabilidad del instrumento fue estimada mediante un doble procedimiento: a) acuerdo inter-jueces, y b) cálculo de la fiabilidad del acuerdo.

a) Acuerdo inter-jueces:

Para realizar el análisis cualitativo de los resultados de la EMODIANA, se estudió el grado de acuerdo entre tres jueces para cada ítem de valoración subjetiva de la muestra. Para ello, se construyó una tabla con 89 casos válidos y se estableció el sistema de categorías, con el propósito de analizar las sesiones de juegos motores y videojuegos.

Los ítems de valoración subjetiva o justificaciones dadas por los niños y niñas, fueron clasificados como internos o externos a la intervención. La dimensión interna incluye las justificaciones referidas al sujeto o persona (P) y las relacionadas a la estructura de la actividad (A). La dimensión externa comprende a las justificaciones referidas al contexto (C). Cada juez debía identificar las categorías que representaran con más precisión la afirmación subjetiva de la justificación de la emoción declarada por cada sujeto. De esta forma se establecieron 7 categorías posibles, las cuales eran mutuamente excluyentes:

1. Persona (P): Cuando la justificación se realiza en primera persona, vinculándose a la justificación de la emoción declarada. Por ejemplo: “porque me reí y me divertí”, “porque tengo miedo de portarme mal como la última vez”

2. Estructura de la actividad (E): Cuando la justificación se realiza sobre elementos o agentes de la intervención específica programada. Por ejemplo: “porque hoy voy a hacer ejercicios nuevos y voy a aprender a hacer juegos”, “porque los juegos eran muy divertidos”.

3. Contexto (C): Cuando la justificación se realiza sobre aspectos del entorno del niño y de la actividad programada. Por ejemplo: “Porque hoy es el cumpleaños de mi padre”, “porque tengo puente y ahora vienen las navidades”.

4. Persona y estructura de la actividad (PE): Cuando se dan de forma conjunta las justificaciones relacionadas con las categorías P y E. Por ejemplo: “porque me siento bien al hacer el ejercicio que he hecho”, “porque me gustan mucho los juegos pero me he portado mal”.

5. Persona y contexto (PC): Cuando se dan de forma conjunta las justificaciones relacionadas con las categorías P y C. Por ejemplo: “porque hoy tuve un buen en clase, en mi casa y aquí”, “porque estoy satisfecho porque ya hice un examen”.

6. Estructura de la actividad y contexto (EC): Cuando se dan de forma conjunta las justificaciones relacionadas con las categorías E y C. Por ejemplo: “...” OJO!!!

7. Persona, estructura de la actividad y contexto (PEC): Cuando se dan de forma conjunta las justificaciones relacionadas con las tres categorías P, E y C. Por ejemplo: “porque ya no vamos a venir más, solo una semana más”, “porque me siento satisfecho cada vez que vengo aquí, y además hoy tuve excursión”.

Sobre este sistema de categorías, los jueces llegaron a los siguientes acuerdos:

-Dentro de la categoría “estructura de la actividad” se acordó incluir a los agentes que forman parte de la actividad, es decir, los compañeros y profesores.

-Se acordó no inferir emociones de la oración sino limitarse a realizar la categorización atendiendo solo al texto.

-Se acordó que el contexto es todo aquello ajeno a la estructura de la actividad.

El procedimiento del acuerdo inter-jueces se llevó a cabo en diferentes fases, realizándose 3 iteraciones (pruebas sucesivas iterativas):

-Pruebas intra-sujeto: se realizan dos medidas propias, dejando pasar una semana entre cada una de ellas para la fiabilidad intra-sujeto.

-Prueba inter-jueces: valoración conjunta de los tres jueces y realización de acuerdos previos.

Con este procedimiento se buscaba evaluar la fiabilidad intrínseca de cada ítem, así como la fiabilidad de los observadores. Si el acuerdo inter-jueces es significativo, podremos concluir la objetividad intrínseca de los ítems.

b) Cálculo de la fiabilidad del acuerdo: Kappa de Fleiss

Como hemos mencionado en la sección introductoria, la Kappa de Cohen es una medida de acuerdo entre dos evaluadores, utilizado para datos categóricos o cuando la variable de respuesta se clasifica mediante una escala nominal o multinominal, que corrige el azar (Losada y Arnau, 2000). Cohen (1960) formuló el cálculo para la obtención de la probabilidad en la concordancia entre dos codificadores, conocida como coeficiente Kappa de Cohen y definida por el símbolo k . Los coeficientes Kappa (k) registran valores que van desde 0 a 1, siendo 0 el valor donde hay mayor desacuerdo entre investigadores/observadores y 1 el punto donde encontramos mayor acuerdo.

Fleiss (1981) generalizó la aplicación del índice Kappa de Cohen para medir el acuerdo entre más de dos codificadores u observadores para datos de escala nominal y ordinal (Torres et al, 2009). Dado que en nuestro estudio tenemos tres evaluadores, y dada la dificultad del fenómeno de justificación de emociones en niños y niñas, procedemos a una validación que contenga

opiniones de expertos relacionadas a distintos campos disciplinares (psicología, juego motor y videojuegos). Por ello, utilizaremos la Kappa de Fleiss para el cálculo de fiabilidad del instrumento, siguiendo las propuestas de diferentes autores (Fleiss, 1981).

El cálculo de este estadístico se realiza de la siguiente manera:

Sea n = el número de ítems, k = el número de categorías de evaluación y m = el número de jueces de cada ítem. En nuestro caso, en la Kappa de Fleiss, tres jueces evalúan cada ítem, tenemos 89 ítems y 7 categorías, entonces tenemos que: $n = 89$, $k = 7$ y $m = 3$.

Por cada ítem $i = 1, 2, \dots, n$, y cada categoría de evaluación $j = 1, 2, \dots, k$ jueces x_{ij} = el número de jueces que asignan la categoría j a i . De esta forma tendremos:

$$0 \leq x_{ij} \leq m \quad \sum_{j=1}^k x_{ij} = m \quad \sum_{i=1}^n \sum_{j=1}^k x_{ij} = mn$$

La proporción de jueces que están de acuerdo en su evaluación sobre el ítem i estará dada por:

$$p_i = \frac{\sum_{j=1}^k C(x_{ij}, 2)}{C(m, 2)} = \frac{\sum_{j=1}^k x_{ij}(x_{ij} - 1)}{m(m - 1)} = \frac{\sum_{j=1}^k x_{ij}^2 - \sum_{j=1}^k x_{ij}}{m(m - 1)} = \frac{\sum_{j=1}^k x_{ij}^2 - m}{m(m - 1)}$$

Por tanto, la media de la P_i es:

$$p_a = \bar{p} = \frac{1}{n} \sum_{i=1}^n p_i = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^k x_{ij}^2 - m}{m(m - 1)} = \frac{1}{mn(m - 1)} \left[\sum_{i=1}^n \sum_{j=1}^k x_{ij}^2 - mn \right]$$

Para el error utilizaremos la siguiente medida:

$$p_\varepsilon = \sum_{j=1}^k q_j^2$$

Dónde:

$$q_j = \frac{1}{nm} \sum_{i=1}^n x_{ij}$$

La Kappa de Fleiss se calcula y define como:

$$\kappa = \frac{p_a - p_\varepsilon}{1 - p_\varepsilon}$$

También se puede definir la Kappa de Fleiss para la categoría j -ésima de la siguiente forma:

$$\kappa_j = 1 - \frac{\sum_{i=1}^n x_{ij}(m - x_{ij})}{mn(m - 1)q_j(1 - q_j)}$$

Luego de aplicar las fórmulas descritas anteriormente para nuestro estudio, los valores que obtuvimos fueron los siguientes: $m = 3$; $n = 89$; $P_a = 0,936$; $P_\varepsilon = 0,338$ y $\kappa = 0,903$. La fuente de datos para este cálculo se describe en la tabla 3.

Para interpretar los valores de los coeficientes Kappa κ obtenidos, podemos utilizar las interpretaciones propuestas por Fleiss (1981) o una versión más amplia, como las propuestas por Altman (1991). Según Fleiss (1981), los coeficientes pueden ser: Regulares (0.40 a 0.60), Buenos (0.61 a 0.75) o Excelentes (superiores a 0.75). Según Altman (1991) los coeficientes Kappa pueden ser Pobres 0 a 0.20), Débiles (0.21 a 0.40), Moderados (0.41 a 0.60), Buenos (0.61 a 0.80) o Muy buenos (0.81 a 1.00).

Por otra parte, para contrastar los resultados obtenidos según el cálculo anterior, utilizamos un método alternativo a la Kappa de Fleiss propuesta por Randolph (2005) 1, obteniendo un valor de $\kappa = 0,936$. Como se puede observar, en ambos casos el índice de fiabilidad κ es mayor que 0,9.

IV. CONCLUSIONES

Los resultados del análisis de las categorías correspondientes a las justificaciones cualitativas observadas muestran que, en las sesiones de juego motor, existe una fuerte relación de las emociones observadas con la estructura de la actividad y la persona (PE), siendo la categoría más observada y ninguna entre la persona y el contexto (PC), la cual no obtuvo ningún registro observacional. Asimismo, se constata una relación directa entre las emociones positivas con la actividad basada en juegos (juego motor y videojuego activo), ya que se observa que las mismas aumentan levemente a la salida de la sesión (95%) y desaparecen casi por completo las emociones negativas (1,64%), permaneciendo alguna emoción neutral, que en el caso del juego podríamos considerarla positiva (3,28% de sorpresa). Podemos decir entonces que, en la experimentación basada en juegos motores y videojuegos activos, la lógica interna (persona y actividad) tiene mayor prevalencia que la lógica externa sobre las emociones positivas generadas y que es posible entonces favorecer ambientes positivos para su desarrollo.

Nuestro instrumento permite medir las emociones percibidas de los niños y niñas frente al reconocimiento de expresiones faciales agregando emociones justificativas (o declarativas) sobre sus percepciones. Si como parece, hay un componente innato en el reconocimiento de las expresiones faciales, culturalmente los niños han aprendido a identificar las emociones. Por eso, cabe confirmar el breve discurso con el cual los niños justifican su estado percibido para las emociones señaladas intuitivamente.

Debido a la naturaleza de la práctica (videojuego y juego motor, videojuego activo-cognitivo, práctica motriz) es esperable la declaración positiva en razón que la naturaleza del juego (juego motor y videojuego) es placentera. De ahí, la distribución encontrada muy centrada en la alegría. Pero esta cuestión se reconduce con la declaración de los niños, quienes matizan la reubicación del carácter positivo y establecen de manera añadida el vínculo entre lo percibido y lo declarado.

Por otra parte, se ha trabajado con un criterio de exhaustividad (categorías), en cuanto a las emociones nuevas que se agregaron a un instrumento de medición emocional ya

validado (Premo) y el problema del acuerdo científico sobre las categorías para las emociones relacionadas con los juegos motores y los videojuegos y las emociones tienen que buscarse en el ámbito de la persona. Las dimensiones y categorías han servido para nuestra coherencia procedimental, mostrando la validez interna y de criterio.

Para comprobar la validez del instrumento y el índice de acuerdo entre los observadores aplicamos la Kappa de Fleiss obteniendo un valor de $k=0,903$ y luego lo contrastamos con la Kappa de Randolph (2008), obteniendo un valor de $k=0,93$. Como podemos observar, ambos índices fueron superiores a 0,9, demostrando así la fiabilidad del instrumento y de la concordancia inter-observacional.

Por ello, concluimos que el instrumento con el procedimiento descrito, es válido para realizar el procedimiento observacional para emociones propias y contenidos de análisis interdisciplinares.

AGRADECIMIENTOS

Este trabajo ha sido financiado por el Proyecto PROVITAO OBE05 de la Fundación Caja Canarias. Asimismo, los autores quieren agradecer a todo el equipo de investigación del proyecto y a los niños y niñas participantes en el estudio.

REFERENCIAS

- [1] Altman, D.G. (1991). *Practical statistics for medical research*. New York: Chapman and Hall.
- [2] Bradley M. & Lang P. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, vol. 25, 1994, pp. 49-59.
- [3] Capozzoli, M., McSweeney, L. & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics*, 27(1), 3-23.
- [4] Cerda L. J. & Villarroel del P. L. (2008). Evaluación de la concordancia inter-observador en investigación pediátrica: Coeficiente de Kappa. *Revista Chilena de Pediatría*, 79 (1): 54-58.
- [5] Desmet P. (2005). *Measuring emotion: development and application of an instrument to measure emotional responses to products*. Funology. Kluwer Academic Publishers, 2005, p. 123.
- [6] Ekman P. & Friesen W. (1975). *Unmasking the face: A guide to recognizing emotions from facial clues*, Prentice-Hall Englewood Cliffs, NJ.
- [7] Fleiss, J. L. & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, vol. 33 pp. 613-619
- [8] Fleiss, J.L. (1981). *Statistical methods for rates and proportions*. New York: John Wiley and Sons.
- [9] González C., Navarro V. & Cairós M. (2013). EMODIANA: Un instrumento para la evaluación subjetiva de emociones en niños y niñas. En *Actas de Interacción 2013*. 17-20 Septiembre 2013. Madrid.
- [10] Holsti, O. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.
- [11] Kaiser S. & Wehrle T. (2001). Facial expressions as indicators of appraisal processes. *Appraisal processes in emotion: Theory, methods, research*, 2001, pp. 285-300.
- [12] Lewis M., Haviland-Jones J, & Barrett L. (2008). *Handbook of emotions*, The Guilford Press.
- [13] Losada, J. L. & Arnau, J. (2000). Fiabilidad entre observadores con datos categóricos mediante el Anova. *Psicothema*, 12 (Supl. 2), 335-339.
- [14] Moggridge, B., & Atkinson, B. (2007). *Designing interactions (Vol. 14)*. Cambridge: MIT press.
- [15] Padilla-Zea N., González-Sánchez J., Gutiérrez-Vela F., Abad-Arroz A. & López-Arco, R. (2012). Evaluación de Emociones en Videojuegos Educativos. El caso particular de los Niños. En *Actas del Congreso Interacción 2012*. Elche. España. 3-5 Octubre.
- [16] Picard, R. & Daily, S.B. (2001). Evaluating affective interactions: Alternatives to asking what users feel. Presented at the 2005 CHI Workshop 'Evaluating Affective Interfaces'.
- [17] Pons, F. (2003). Individual differences in children's emotion understanding: effects of age and language. *Scandinavian Journal of Psychology*, 44, 347-353.
- [18] Randolph J. (2005) Free-marginal multirater kappa (multirater κ free): an alternative to Fleiss' fixed-Marginal multirater kappa. Paper presented at the Joensuu Learning and Instruction Symposium, Joensuu, Finland
- [19] Schneider K & Josephs I. (1991). The expressive and communicative functions of preschool children's smiles in an achievement-situation. *Journal of Nonverbal Behavior*, vol. 15, 1991, pp. 185-198.
- [20] Torres, J.J. & Perera, V.H. (2009). Cálculo de la fiabilidad y concordancia entre codificadores de un sistema de categorías para el estudio del foro online en e-Learning. *Revista de Investigación Educativa*, 27, p. 89-103