

MEDICION INFORMAL DEL P-VALOR: UN ESTUDIO EXPLORATORIO CON ESTUDIANTES DE BACHILLERATO

Víctor N. García¹, Ernesto Sánchez²

Departamento de Matemática Educativa, CINVESTAV-IPN

nozairg@hotmail.com¹

esanchez0155@gmail.com²

En los años recientes se ha reconocido la importancia de desarrollar el razonamiento inferencial informal (RII), antes de aprender los conceptos estadísticos formales. No obstante, hay poca investigación sobre su desarrollo en el aula en el nivel bachillerato. En situaciones informales, una de las dificultades que sobresale es la de medir el p-valor de un estadístico (probabilidad de obtener el estadístico o más extremo si la hipótesis nula fuera verdadera). El presente trabajo es un avance de investigación, en el cuál se analiza el razonamiento de estudiantes de bachillerato (15-18 años) para medir el p-valor de un estadístico al usar una simulación computarizada que crea una distribución muestral empírica. Se encontró que varios estudiantes miden adecuadamente el p-valor de manera informal mediante la simulación, lo que representa un cambio significativo en el RII.

Palabras claves: Razonamiento inferencial informal, prueba de significación, distribución muestral, p-valor.

Introducción

La literatura sobre inferencia estadística muestra que es un tema difícil de aprender (Castro-Sotos, Vanhoof, Van den Noorgate, & Onghena, 2007). Los resultados estudios empíricos informan que estudiantes, profesores e incluso profesionales cometen errores conceptuales. Una de las razones proviene del hecho de que el tratamiento de la inferencia en los cursos de estadística del bachillerato y primer curso universitario, se enfoca hacia el aprendizaje de procedimientos rutinarios y la reproducción de la exposición formal, sin ofrecer a los estudiantes la oportunidad de discutir y entender las ideas fundamentales subyacentes en los procedimientos inferenciales antes de formalizarlas con los conceptos matemáticos correspondientes.

Esto ha motivado el interés por estudiar la Inferencia Estadística Informal (IEI) y el Razonamiento Inferencial Informal (RII), como lo confirman publicaciones de números especiales (Por ejemplo, Makar & Ben-Zvi, 2011), con los objetivos de: 1) descubrir y describir formas como los estudiantes desarrollan ideas fundamentales de la inferencia estadística sin utilizar el aparato matemático que las fundamenta, y 2) crear un repertorio de problemas y actividades que jueguen un papel de antecedente en el aprendizaje de los estudiantes sobre el cual puedan construir los conocimientos formales de la inferencia estadística. En este contexto, investigadores han visto a la

tecnología, en particular a las simulaciones, como una opción para llevar a cabo los objetivos anteriores. Sin embargo, se han realizado pocos estudios con alumnos de bachillerato (15-18 años), nivel en el que se centra nuestra investigación.

Los estudiantes presentan una falta de comprensión del concepto de distribución muestral y, con ello, dificultades para medir el p-valor. Por tanto, realizan inferencias estadísticas inadecuadas (Chance, delMas & Garfield, 2004). Al medir el p-valor del estadístico de forma intuitiva los estudiantes lo hacen erróneamente. El presente estudio pretende mostrar cómo una simulación de una distribución muestral empírica tiene el potencial para desarrollar informalmente la medición del p-valor. Las preguntas de investigación que ayudaran a lograr el objetivo son: ¿Cómo se caracteriza el razonamiento para hacer inferencias estadísticas informales con el uso de la simulación que llevan a cabo estudiantes de bachillerato? ¿Qué errores y dificultades persisten en el razonamiento para hacer inferencias estadísticas informales con el uso de la simulación?

ANTECEDENTES

Varios trabajos publicados en los últimos años aluden a los conceptos de IEI y RII (por ejemplo, Makar & Ben-Zvi, 2011); sin embargo, todavía no hay consenso acerca de lo que significan estos dos términos exactamente. En un intento de combinar las distintas perspectivas, Zieffler, Garfield, delMas, y Reading (2008) definen RII como “la forma en que los estudiantes usan sus conocimientos informales de estadística para crear argumentos basados en muestras observadas para sustentar las inferencias sobre la población desconocida” (p. 44). Estos autores, también proponen un marco conceptual para caracterizar y desarrollar el RII. Las investigaciones sobre RII se pueden dividir en dos enfoques diferentes, a saber, uno que contiene estudios sobre la naturaleza del RII y otro que incluye estudios que se refieren al desarrollo del RII. Analizaremos trabajos con ambos enfoques, comenzando con los estudios de la naturaleza del RII.

Trabajos relacionados con las ideas detrás del RII varias décadas atrás son los de Khaneman y Tversky (1982), donde estudian las intuiciones estadísticas de las personas en situaciones de incertidumbre, encontrando que las personas tienden a usar heurísticas para medir intuitivamente probabilidades que generalmente son inadecuadas y llevan a tomar decisiones erróneas. El marco elaborado por Zeiffler et al. (2008) propone diseñar tareas que permitan explorar y desarrollar el RII, estas tareas deben de provocar en los alumnos tres tipos de actividades: a) hacer juicios o predicciones sobre poblaciones con base en una muestra, b) recuperar, utilizar e integrar conocimientos previos disponibles y pertinentes y c) elaborar argumentos para apoyar los juicios o predicciones.

En García-Ríos (2013) se informa que los estudiantes tienen dificultades para evaluar la significatividad del estadístico de la muestra adecuadamente, debido a dos

posibles causas: a) Razonan de forma determinista al medir la significatividad del estadístico de la muestra, en el sentido de que el estadístico debe coincidir exactamente con el parámetro poblacional. b) Comparan el estadístico con un modelo probabilístico inapropiado de la población (hipótesis nula), creado por el estudiante con base en sus conocimientos. Además, utilizan muchos prejuicios y creencias en sus argumentos y no reflejan grados de incertidumbre en sus inferencias. García y Sánchez (2014) han observado que la concepción de Fisher de las pruebas de significación es algo natural para los estudiantes, pues establecen una hipótesis nula (modelo personal de la población) para comparar la muestra y medir intuitivamente su significatividad (aunque de manera inapropiada). Los autores anteriores consideran que es importante que los estudiantes trabajen con los datos y ubiquen en un segundo plano sus conocimientos informales personales sin descartarlos por completo, y concluyen que para desarrollar inferencias apropiadas antes de su formalización, es crucial que los estudiantes cuenten con un método informal para determinar un criterio numérico para saber cuándo rechazar o no la hipótesis. En consecuencia se considera que la simulación es un recurso que suministra dicho método.

Con respecto a los estudios para desarrollar el RII, Rossman (2008) sugiere una introducción informal siguiendo los pasos siguientes: (a) Partir de una hipótesis dada acerca de los datos. (b) Uso de la simulación para concluir que los datos observados son muy poco probables si el modelo fuera cierto (cálculo intuitivo de un p-valor). (c) Rechazar la hipótesis inicial basado en el p-valor muy pequeño. Este proceso de razonamiento, parece natural para los estudiantes, y de hecho sigue la concepción de Fisher de pruebas de significación. Pfannkuch (2005) considera que el enfoque de “re-muestreo” tiene el potencial de hacer una fuerte conexión entre la probabilidad y el análisis de datos, pues está más en sintonía con una conceptualización de un P-valor, y Fathom permite este enfoque. Hofmann, Maxara, Meyfarth y Prommel (2014) ven en Fathom un gran potencial didáctico para la enseñanza y aprendizaje de la estadística con simulaciones. Ellos proponen que para tener éxito en la enseñanza con Fathom es necesario diseñar tareas adecuadas, contar con fases de familiarización de la tarea, de presentación, de trabajo de los estudiantes y de discusión.

Jacob y Doerr (2013) investigan el razonamiento estadístico informal de siete pares de estudiantes (16-18 años). Se encontró que, aunque en general los estudiantes tenían un buen conocimiento de la distribución muestral, la mayoría de ellos no comparan la muestra con la distribución muestral. Estos autores sugieren elaborar actividades que den la oportunidad de recoger una muestra de datos y compararlo con la distribución muestral en relación con las probabilidades asociadas, para ayudar a entender cómo una sola muestra se puede utilizar para hacer una conclusión acerca de una población.

MARCO CONCEPTUAL

En este trabajo, se entiende por marco conceptual a una red de conceptos o categorías relacionados entre sí que en conjunto proporcionan una comprensión global de un fenómeno o fenómenos (Jabareen, 2009). Además, es posible construir un marco conceptual emergente de los datos del estudio mediante la metodología de la teoría fundamentada (Birks & Mills, 2011).

Pruebas de significación

Dentro de la inferencia estadística hay dos concepciones sobre los contrastes de hipótesis: (a) Las pruebas de significación, que fueron introducidas por Fisher y (b) Los contrastes como reglas de decisión entre dos hipótesis, que fue la concepción de Neyman y Pearson. La enseñanza ignora estas diferencias y presenta los contrastes de hipótesis como si se tratase de una única metodología (Batanero, 2000). Se empleará la concepción de las pruebas de significación de Fisher para introducir el contraste de hipótesis ya que se tiene la hipótesis de ser pertinente (García y Sánchez, 2014).

Razonamiento inferencial informal

La inferencia estadística informal (IEI) es una generalización probabilística (no determinista) de los patrones que son revelados por los datos disponibles, y esta generalización es el producto final de un razonamiento inferencial informal (RII) (Makar, Bakker & Ben-Zvi, 2011). Una IEI se representa mediante un enunciado, mientras que el RII es el proceso mediante el cual se descubren y establecen dichos enunciados. El RII es la forma en que los estudiantes usan sus conocimientos para hacer y sustentar inferencias estadísticas sobre una población desconocida basadas en muestras observadas y sin utilizar los métodos o técnicas formales de la estadística inferencial.

La tecnología en la enseñanza de la estadística

La tecnología ha ampliado las técnicas gráficas y de visualización para proporcionar nuevas y poderosas formas de ayudar a los estudiantes en el análisis de datos, lo que les permite centrarse en la interpretación de los resultados, en la comprensión y utilidad de conceptos e ideas estadísticas fundamentales, y no en la mecánica computacional. Erickson (2006) cree que los conceptos difíciles y el proceso del contraste de hipótesis se hacen más visibles y comprensibles por medio de la simulación con Fathom, ya que la velocidad de la computadora hace que sea posible hacer mucho más ensayos y permite centrarse en la distribución muestral, ayudando a elevar el nivel de abstracción. Además el estudiante puede variar parámetros y describir y explicar el comportamiento que observan en lugar de confiar exclusivamente en las discusiones teóricas de probabilidad, pues estas, a menudo son contrarias a la intuición de los estudiantes (delMas, Garfield & Chance, 1999). El entorno visual, interactivo y dinámico

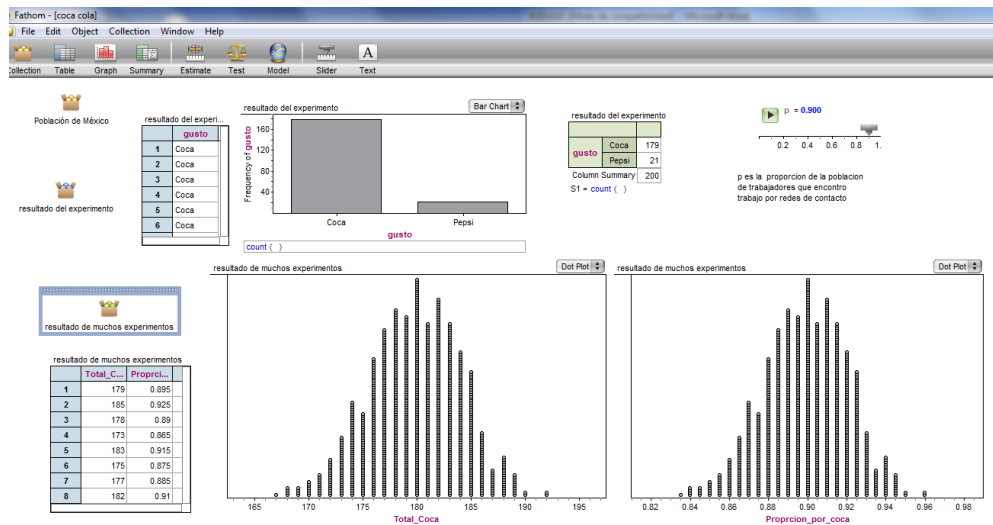
de la tecnología ayuda a lograr estos objetivos (Chance, Ben-Zvi, Garfield & Medina, 2007).

Para que la tecnología tenga impacto es importante enfocarse en buscar formas eficientes de usarla en el salón de clase. Es importante el empleo del aprendizaje colaborativo y facilitar la interacción y accesibilidad a la tecnología, manteniendo el enfoque en los conceptos estadísticos en lugar de la herramienta (Biehler, Ben-Zvi, Bakker & Makar, 2013). La simulación está creada en Fathom (Finzer, 2014); un software dinámico y animado creado para la enseñanza y aprendizaje de la Estadística. Permite realizar procesos estadísticos como crear tablas, gráficas, muestreo, definir y calcular estadísticos y crear simulaciones animadas. Una característica importante es que un conjunto de datos se puede representar al mismo tiempo en una tabla y en una gráfica y al cambiar en alguna de ellas un dato, la otra representación se actualiza automáticamente, así como todas las funciones y procedimientos que hayan sido definidos con el conjunto inicial de datos.

Metodología

La recolección de datos se llevó a cabo mediante un problema de prueba de hipótesis de proporciones; se aplicó a 19 parejas de estudiantes de nivel bachillerato (edades entre 16 y 17 años) en una escuela pública. Los estudiantes respondieron en el ordenador con la posibilidad de usar la simulación.

Figura 1. Simulación interactiva para la prueba de significación.



Antes de aplicar el problema se explicó el funcionamiento de la simulación mediante un problema introductorio. Además se contó con hojas de trabajo orientadas a los procesos y se trabajó en un aprendizaje cooperativo con 2 estudiantes por computadora. La sesión fue de una hora y la llevo a cabo uno de los autores de la

comunicación. Después de la sesión introductoria se les presentó el problema y se dio un tiempo de dos horas para responder las preguntas. Durante el cuestionario los estudiantes podían comentar sus dudas con el profesor y discutir entre ellos.

El problema utilizado para la recolección de datos es el siguiente: Propaganda de la Coca Cola presume que el 90% de la población de México prefiere su refresco que Pepsi. Para comprobarlo se les dieron dos vasos de refresco (uno con Coca y otro con Pepsi) a 200 personas escogidas al azar y decidían cuál le gustaba más. De los 200 participantes 188 personas prefirieron Coca Cola. ¿El resultado del experimento es suficiente para decir que el 90% de la población de México prefiere Coca Cola? ¿Por qué?

La distribución muestral (figura 1) se construye simulando 1000 encuestas a 200 personas tomadas de una población hipotética con parámetro 0.9, dicho parámetro podía ser manipulado por un deslizador (parte superior derecha de la figura 1). La última muestra simulada se representa en una gráfica de barras y en una tabla (parte superior de la figura 1). Por medio de la distribución muestral empírica se puede medir la probabilidad de los datos observados con el método frecuencial (empírico), esto es, el cálculo informal de un p-valor mediante frecuencias (Rossman, 2008). La distribución muestral se exhibe en una tabla y en dos gráficas de puntos (parte inferior de la figura 1), tanto en valores absolutos como en proporciones; cada punto de la gráfica describe el total de participantes que prefirieron Coca Cola, ya sea en forma absoluta o en proporción.

El diseño de la simulación pretende ser accesible y facilitar la interacción con los estudiantes, tratando de mantenerlos enfocados en los conceptos estadísticos. La animación hace que el estudiante observe todo el proceso de re-muestreo simultáneamente; obtener una muestra, obtener el valor del estadístico de la muestra, graficar.

Resultados

Tabla 1. Categorías de análisis

Medición del p-valor	Argumentación	Conclusión
Simulación 12 (de 19 parejas)	Muestra rara	Rechazar hipótesis
	Muestra común	No rechazar hipótesis
Datos 4 (de 19 parejas)	Proporción diferente a la hipótesis (mayor o menor)	Rechazar hipótesis
		No rechazar hipótesis
Creencias 3 (de 19 parejas)	Población y publicidad	Rechazar hipótesis
		No rechazar hipótesis

de razonamiento son considerados como intuitivos (García & Sánchez, 2015) y es necesario desarrollar la noción de distribución muestral de una manera informal para alcanzar un nivel mayor en el RII. Estas respuestas se clasificaron en “Datos” y “Creencias”.

Datos. Dentro de esta categoría se ubican las respuestas donde el razonamiento se basa en los datos de las muestras y no consideran que hay variación del estadístico de muestra a muestra, así, si la proporción de la muestra es mayor que la hipótesis entonces concluyen que es más del 90% y rechazan la hipótesis. Por ejemplo E8 responden “porque el 90% de 200 es 180 por lo tanto la mayoría de las personas prefieren coca porque esta superior que la Pepsi según los participante de la encuesta”.

Creencias. En esta categoría se clasificaron las respuestas que no consideran los datos para argumentar su conclusión. Por ejemplo E2 responde “coca cola es una empresa mundial y por lo cual tiene mayor publicidad, mayor calidad probablemente”. Una respuesta por demás interesante es la dada por E6 “Porque los resultados son al azar, entonces alguna otra parte de la población que no fue tomada en la encuesta prefiere Pepsi, así que los porcentajes cambian y no se puede saber con seguridad”. Se tiene presente la variación, sin embargo esta en un grado excesivo, al no considerar que una muestra pueda ser representativa.

Discusión

Como parte de la discusión se dará respuesta a las preguntas de investigación planteadas en la introducción y comentarios sobre los resultados encontrados.

¿Cuál es el razonamiento para hacer inferencias estadísticas informales con el uso de la simulación que llevan a cabo estudiantes de bachillerato? Los resultados dan indicios de que varios estudiantes parecen comprender el proceso de construcción de la distribución muestral por medio de la simulación y que pueden determinar el p-valor del estadístico de la muestra (si es rara o no) a partir de la distribución muestral empírica. Este resultado es muy importante pues se ha documentado que los estudiantes tienen dificultad para medir el p-valor de forma intuitiva: si el estadístico es diferente al parámetro entonces se rechaza la hipótesis, sin importar la diferencia (García & Sánchez, 2015). Otros estudiantes requieren que la muestra esté en los extremos del parámetro para rechazar la hipótesis (García-Ríos, 2013). De aquí que sea conveniente que desarrollen un método informal para medir el p-valor, donde la simulación junto con más actividades bien diseñadas tienen el potencial para desarrollar el RII.

Se han observado que persisten las limitaciones en la forma intuitiva de razonar de los estudiantes: 1) Buscan la respuesta en los datos pero sin tener un modelo de probabilidad pues solo se enfocan en el estadístico de la muestra y lo comparan con el

parámetro de la población (García & Sánchez, 2015). 2) Dan argumentos basados en creencias personales, sin llevar a cabo procedimientos que utilicen los datos.

¿Qué errores y dificultades persisten en el razonamiento para hacer inferencias estadísticas informales con el uso de la simulación? No toman en cuenta la distribución muestral empírica ya sea por basarse en creencias para explicar los datos o solo enfocarse en el estadístico de la muestra sin un modelo de probabilidad. También es posible que no se comprendiera el proceso de construcción de la distribución muestral empírica, provocando la falta de percepción de la variación natural de las muestras y su representatividad.

Un dato interesante y que no se analiza en esta comunicación es el hecho de que varios estudiantes parecen comprender el proceso de construcción de la distribución muestral mediante la simulación y que pueden determinar el p-valor del estadístico, pero llegan a diferentes conclusiones; estudiantes que encontraron una muestra normal rechazaron la hipótesis cuando no debían. Esta inconsistencia muestra que no son suficientes los resultados de la simulación.

Un posible camino para atacar estas dificultades es diseñar actividades con variantes en los parámetros y discusiones, para hacer notar por qué algunos argumentos no son adecuados. Lo presentado aquí sólo es una actividad de varias que se han planeado para investigar el tema. Además se ha centrado en un solo componente del proceso del contraste de hipótesis. Sin embargo los resultados obtenidos aportan información para refinar las actividades previstas y para construir un modelo de desarrollo del RII.

Referencias

- Batanero, C. (2000). Controversies around significance tests. *Mathematical Thinking and Learning*, 2(1-2), 75-98.
- Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. En A. Bishop, K. Clement, C. Keitel, J. Kilpatrick, & A. Y. L. Leung (Eds.), *Third international handbook on mathematics education* (pp. 643-689). New York: Springer.
- Birks M. & Mills, J. (2011). *Grounded theory: A practical guide*. California: Sage.
- Castro-Sotos, A. E., Vanhoof, S., Van den Noorgate, W., Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2, 98-113.
- Chance, B. L., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning. *Technology Innovations in Statistics Education*, 1(1).
- Chance, B., delMas, R. C., & Garfield, J. (2004). Reasoning about sampling distributions. En D. Ben-Zvi & J. Garfield (Eds.). *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Amsterdam: Kluwer Academic Publishers.

- delMas, R., Garfield, J., & Chance, B. (1999). A model of classroom research in action: Developing simulation activities to improve student's statistical reasoning. *Journal of Statistics Education*, 7(3).
- Erickson, T. (2006). *Using simulation to learn about inferences*. *International Conference on Teaching Statistics*, 7, 1-6.
- Finzer, W. (2014). *Fathom Dynamic Data Software (Version 2.2)* [Software]. Disponible: <http://fathom.concord.org/>.
- García-Ríos, N. (2013). Inferencias estadísticas informales en estudiantes mexicanos. En J. M. Contreras, G. R. Cañadas, M. M. Gea y P. Arteaga (Eds.), *Actas de las Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria* (pp. 343-357). Granada, 2013.
- García, V. N., & Sánchez, E. A. (2014). Razonamiento inferencial informal: el caso de la prueba de significación con estudiantes de bachillerato. En M. T. González, M. Codes, D. Arnau y T. Ortega (Eds.), *Investigación en Educación Matemática XVIII* (pp. 345-354). Salamanca: SEIEM.
- García, V. N., & Sánchez, E. A. (2015). Dificultades en el razonamiento inferencial intuitivo. En J. M. Contreras, C. Batanero, J. D. Godino, G.R. Cañadas, P. Arteaga, E. Molina, M.M. Gea y M.M. López (Eds.), *Didáctica de la Estadística, Probabilidad y Combinatoria*, 2 (pp. 207-214). Granada, 2015.
- Hofmann, T., Maxara, C., Meyfarth, T., & Prömmel, A. (2014). Using the Software FATHOM for learning and teaching statistics in Germany –A review on the research activities of Rolf Biehler's working group over the past ten years. En T. Wassong, D. Frischmeier, P. Fischer, R. Hochmuth & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen –Using Tools for Learning Mathematics and Statistics* (pp. 283-304). Wiesbaden: Springer.
- Jabareen, Y. (2009). Building conceptual framework: Philosophy, definitions and procedure. *International Journal of Qualitative Method*, 8(4), 49-62.
- Jacob, B., & Doerr, H. (2013, February). Students' informal inferential reasoning when working with the sampling distribution. Paper presented in Working Group 5 (A. Bakker, Chair): *The Eighth Congress of European Research in Mathematics Education (CERME8)*, Manavgat-Side, Antalya: Turkey
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Makar, K., Bakker, A. & Ben-Zvi, D. (2011). The Reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, 13(1-2), 152-173.
- Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning*, 13(1-2), 1-4.
- Pfannkuch, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection? En G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 267-294). New York: Springer.
- Rossmann, A. (2008). Reasoning about informal statistical inference: one statistician's view. *Statistics Education Research Journal*, 7(2), 5-19.
- Zieffler, A., Garfield, J., delMas, R. & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40-58.