



Universidad de León

TESIS DOCTORAL

Curso 2019/2020

**ANÁLISIS Y EVALUACIÓN DEL USO DE LA  
SUPERCOMPUTACIÓN EN LA MEJORA DEL  
DESEMPEÑO FORMATIVO**

ANALYSIS AND EVALUATION OF SUPERCOMPUTING FOR TRAINING  
PERFORMANCE IMPROVEMENT

Realizado por el Alumno D. Álvaro Fernández González

Dirigido por la Profesora Dra. Dña. Camino Fernández Llamas, siendo el tutor el Profesor  
Dr. D. Vicente Matellán Olivera.

León, 30 de junio de 2020



*«Un niño puede enseñar tres cosas a un adulto: a ponerse contento sin motivo, a estar siempre ocupado con algo y a saber exigir con todas sus fuerzas aquello que desea».*

*Paulo Coelho*



## AGRADECIMIENTOS

En estas primeras líneas, quiero agradecer a todas las personas que, de un modo u otro, me han animado en la realización de esta tesis:

En primer lugar, un agradecimiento especial a mi directora y mi tutor de tesis, Camino y Vicente, por animarme a seguir cuando las fuerzas y los ánimos flaqueaban y por hacerme ver que investigar es algo divertido, que nos permitirá hacer del mundo algo mejor. A José Ángel, mi mentor en la Universidad de León y quien me ha acompañado en este proyecto desde mis primeros pasos. A Miguel Ángel, que ha contribuido al empujón final para subir a la cumbre cuando las fuerzas estaban justas. A Rafael, mi coterráneo, que me transmitió el ánimo de su ilusión en investigar. A mis compañeros de trabajo de SCAYLE donde siempre hemos tenido un ambiente en el que investigar era un componente más de nuestras tareas diarias. En este apartado no puede dejar de mencionar a un antiguo Jefe de esos que marcan el destino: Tácito, alguien que me dio la oportunidad de demostrar que era un persona útil y válida y cuya doctrina aplico siempre en mi día a día.

Por otra parte, me gustaría agradecer a mi familia, mi esposa Sonia y en especial a mis peques (Alejandro, Diego y Sofía) que son mi fuente de inspiración y el motivo por el cual afronto el día a día con ilusión. No puedo dejar tener un recuerdo especial en estos momentos a los seres queridos que ya no están: a mi abuela Ción, que fue mi soporte vital hasta que la vida le duró, a mi abuelo Colás, que se fue siendo yo un niño, pero que me quiso con todas sus fuerzas, a mi abuela Luisa, que siempre me quiso como nadie y a mis tíos, en especial a mi tío Mario, cuya fuerza vital, sus charlas y consejos son hoy en día una fuente de inspiración para mí.

Por último, querría agradecer a mi gran amigo Jesús Callejo y a su equipo del programa de radio “La Escóbulas de la Brújula” (Carlos Canales, Juan Ignacio Cuesta y el resto del grupo) por ser mi soporte en las madrugadas en las que me he quedado desarrollando el trabajo y que han sido mi compañía cuando el cansancio y el sueño llamaban a mi puerta. A todos, gracias, de corazón.



# INDICE

RESUMEN	12
INTRODUCCIÓN	14
1.- OBJETIVOS	17
2.- METODOLOGÍA	19
3.- REVISIÓN HISTÓRICA Y DESAFÍOS FUTUROS EN SUPERCOMPUTACIÓN Y REDES DE COMUNICACIÓN CIENTÍFICA	21
3.1.- INTRODUCCIÓN	22
3.2.- METODOLOGÍA Y PREGUNTAS DE LA INVESTIGACIÓN	23
3.3.- PRINCIPALES HALLAZGOS DE LA INVESTIGACIÓN	25
3.3.1.- P1: ¿Cuándo nació la supercomputación?	25
3.3.2.- P2: ¿Cómo y cuándo se hizo la transición de los usos iniciales de la supercomputación a usos científicos?	29
3.3.3.- P3: ¿Cuáles son los usos actuales de la supercomputación y cuál es el pronóstico para el futuro?	32
3.3.4.- P4: ¿Cómo se han desarrollado las redes de comunicaciones científicas?	43
3.3.5.- P5: ¿Las redes de comunicaciones científicas ayudan a la supercomputación?	48
3.4.- LIMITACIONES DEL ESTUDIO	53
4.- SUPERCOMPUTADORAS PARA MEJORAR EL RENDIMIENTO EN LA EDUCACIÓN SUPERIOR: UNA REVISIÓN DE LA LITERATURA	55
4.1.- INTRODUCCIÓN	57
4.2.- PREGUNTAS DE LA INVESTIGACIÓN	59
4.3.- METODOLOGÍA	60
4.3.1. Estrategia de búsqueda y fuentes de datos.	60
4.3.2.- Criterios de inclusión / exclusión.	62
4.3.3.- Fases de búsqueda y selección de estudios	63
4.3.4.- Evaluación de la calidad del artículo.	64
4.4.- RESULTADOS	67
4.4.1.- P1: ¿Cuáles son los factores considerados que mejoran la formación en Supercomputación?	69
4.4.2.- P2: ¿Cómo puede la formación en supercomputación mejorar los resultados de los estudiantes y/o investigadores?	70
4.4.3.- P3: ¿Qué se sabe actualmente sobre las limitaciones en la formación relacionada con la supercomputación?	72
4.4.4.- P4: ¿Cuáles son las soluciones descritas para resolver las limitaciones descritas en relación con la formación en Supercomputación?	74
4.5.- DISCUSIÓN	76
4.5.1.- Resultados	77
4.5.2.- Limitaciones	78

5.- INTEGRANDO CLÚSTERES DE SUPERCOMPUTACIÓN EN LA EDUCACIÓN: UN ESTUDIO DE CASO EN BIOTECNOLOGÍA	79
5.1.- INTRODUCCIÓN	80
5.2.- MATERIALES Y METODOLOGÍA	85
5.3.- RESULTADOS	86
5.3.1.- Campos de aplicación	88
5.3.2.- Tipo de cursos de contenido computacional relacionados con la supercomputación	90
5.3.3.- Estudio relacional que compara el campo de conocimiento que utiliza la supercomputación y el tipo de capacitación computacional específica	92
5.4.- ESTUDIO DE CASO	95
5.4.1.- Participantes y recopilación de datos	99
5.4.2.- Resultados de la encuesta	100
5.5.- DISCUSIÓN	102
5.5.1.- Supercomputación en el proceso educativo	103
5.5.1.1.- Resultados	104
5.5.1.2.- Limitaciones	104
5.5.2.- Estudio de caso	105
5.5.2.1.- Resultados	105
5.5.2.2.- Limitaciones	107
6.- ANÁLISIS DEL DESARROLLO Y LA VALIDEZ PREDICTIVA DEL CUESTIONARIO DE CAPACITACIÓN RELACIONADO CON LA SUPERCOMPUTACIÓN	108
6.1.- INTRODUCCIÓN	110
6.2.- REVISIÓN DE LITERATURA	111
6.3.- METODOLOGÍA	113
6.3.1.- Pool de ítems.	113
6.3.2.- Validación por expertos.	115
6.3.3.- Prueba piloto	117
6.3.4.- Adaptación de la encuesta modelo TPACK	119
6.3.5.- Estudio final	119
6.4.- ANÁLISIS DE DATOS	120
6.4.1.- Pool de ítems	120
6.4.2.- Comité de Expertos	121
6.4.3.- Prueba piloto	122
6.4.4.- Encuesta final: adaptación de la encuesta modelo TPACK	127
6.4.5.- Estudio final	132
6.5.- RESULTADOS	133
6.5.1.- Análisis factorial	137
6.5.2.- Rotación	142
6.5.2.1.- Rotación Oblimin	143

6.5.2.2.- Rotación varimax	146
6.6.- DISCUSIÓN	150
6.6.1.- Resultados	151
6.6.2.- Limitaciones	152
7.- CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	154
7.1.- CONCLUSIONES DEL TEMA 3	154
7.2.- CONCLUSIONES DEL TEMA 4	156
7.3.- CONCLUSIONES DEL TEMA 5	157
7.4.- CONCLUSIONES DEL TEMA 6	159
8.- PUBLICACIONES Y DIFUSIÓN DE RESULTADOS	162
8.1.- REVISTAS INDEXADAS	162
8.2.- CONGRESOS	162
APÉNDICE I – Artículos de referencia	164
APÉNDICE II – Encuesta: Versión inicial	167
APÉNDICE III – Análisis de los elementos de la encuesta	172
APÉNDICE IV – Evaluación experta: Resultados	174
BIBLIOGRAFÍA	176

## LISTA DE FIGURAS

Figura 3.1: Evolución de la Supercomputación.	29
Figura 3.2: El momento decisivo en la ciencia: El cambio de silicio (SS).	30
Figura 3.3: Proceso continuo de la investigación a la producción	33
Figura 3.4: Cronología de la evolución de ARPANET	43
Figura 3.5: RedCLARA 2013	47
Figura 4.1: Fases de búsqueda y selección	64
Figura 4.2: Número de publicaciones por año	67
Figura 5.1: Diagrama de flujo de un proceso de capacitación relacionado con la supercomputación	88
Figura 5.2: Descripción general del trabajo de los estudiantes en los cursos de supercomputación	98
Figura 5.3: Detalle de herramientas gráficas durante el desarrollo de un curso	99
Figura 6.1: Modelo TPACK	112
Figura 6.2: Representación gráfica de los datos generales de la encuesta	133

## LISTA de TABLAS

Tabla 3.1: Detalle de industrias y sectores que utilizan la supercomputación	35
Tabla 4.1: Puntuación de relevancia de la calidad de los artículos analizados	66
Tabla 4.2: Selección de estudios por tipos de publicación	67
Tabla 4.3: Limitaciones en la formación relacionada con la supercomputación	72
Tabla 4.4: Soluciones a las limitaciones en formación relacionada con la supercomputación	74
Tabla 5.1: Campos de conocimiento que utilizan la supercomputación	89
Tabla 5.2: Cursos computacionales relacionados con la supercomputación	90
Tabla 5.3: Modelos creados según el criterio “campo de conocimiento”	93
Tabla 5.4: Resultados detallados de ANOVA para cada modelo	94
Tabla 5.5: Detalle de la correlación de los modelos	95
Tabla 5.6: Caracterización de los cursos (datos significativos)	96
Tabla 5.7: Perfil de los alumnos	99
Tabla 5.8: Resultados de la encuesta	100
Tabla 6.1: Detalle del Comité de Expertos	115
Tabla 6.2: Sumario de estadísticas de los ítems	122
Tabla 6.3: Estadísticas de fiabilidad	123
Tabla 6.4: Estadística totales de los Ítems	123
Tabla 6.5: Sumario de estadísticas de los ítems	125
Tabla 6.6: Estadísticas de fiabilidad	125
Tabla 6.7: Estadísticas totales de los ítems	126
Tabla 6.8: Variables de entrada y eliminadas <sup>a</sup>	128
Tabla 6.9: Modelos resumidos <sup>e</sup>	129
Tabla 6.10: ANOVA <sup>a</sup>	129
Tabla 6.11 Coeficientes de los modelos generados <sup>a</sup>	130
Tabla 6.12: Cuestionario Final	131
Tabla 6.13: Matrices Anti-imagen	136
Tabla 6.14: Test KMO y Bartlett	137
Tabla 6.15: Comunalidades	138
Tabla 6.16: Varianza total explicada	139
Tabla 6.17: Matriz de Componentes <sup>a</sup>	141
Tabla 6.18: Matriz de patrones <sup>a</sup>	144
Tabla 6.19: Matriz de estructura	145
Tabla 6.20: Componentes de la matriz de correlación	146
Tabla 6.21: Matriz de componentes rotados <sup>a</sup>	147
Tabla 6.22: Component Transformation Matrix	148

## LISTA DE GRÁFICOS

Gráfico 6.1: Gráfico de sedimentación	140
Gráfico 6.2: Gráfico de componentes no rotados	142
Gráfico 6.3: Componentes del espacio rotado	148



## **RESUMEN**

Los recursos de supercomputación son en la actualidad el pilar fundamental para el desarrollo de la investigación en diversos campos. Su impacto se basa en la capacidad de cálculo, que permite realizar simulaciones computacionales que permiten mejorar la precisión de los experimentos. La presente Tesis Doctoral pretende, en primer lugar, realizar un estudio de la evolución de la supercomputación y su aplicación a diversos campos para, posteriormente, estudiar los factores determinantes que permitan analizar los aspectos más relevantes a la hora de estudiar la relación existente entre los estudios de supercomputación con los aspectos pedagógicos, de conocimiento y de contenido, basándose en el modelo TPACK. El estudio se realizó con información procedente de la base de datos de estudiantes del Centro de Supercomputación de Castilla y León (SCAYLE), de la que se obtuvieron 97 participantes. En el estudio se realizó un análisis factorial para comprobar que la estructura de datos obtenida era coherente con el modelo TPACK usado como referencia. Los resultados obtenidos del análisis relacionan las dimensiones tecnológicas con las de conocimiento, pedagógicas y de contenido.

Palabras clave: Supercomputación, TPACK, formación, tecnología.

## **ABSTRACT**

Supercomputing resources are currently the fundamental pillar for the development of research in various fields. Its impact is based on the calculation capacity, which allows computational simulations to be carried out to improve the precision of the experiments. This Doctoral Thesis aims, firstly, to carry out a study of the evolution of supercomputing and its application to various fields. Then, to study the determining factors that allow the most relevant aspects to be analyzed in order to establish the relationship between supercomputing studies and pedagogical, knowledge and content aspects, based on the TPACK model. The study was carried out with data from the student database of the Castilla y León Supercomputing Center (SCAYLE), from which 97 participants. In the study, a factor analysis was performed to verify that the data structure obtained was consistent with the TPACK model used as a reference. The results obtained relate the technological dimensions with those of knowledge, pedagogy and content.

**Keywords:** Supercomputing, TPACK, training, technology

## INTRODUCCIÓN

En la actualidad, la supercomputación es una herramienta de gran importancia en el cambio de paradigma tecnológico, tanto presente como futuro, para dar soporte, a través del crecimiento de la demanda de servicios de cálculo intensivo, a grupos de investigación cuyos resultados no se podrían obtener sin el soporte de un supercomputador. Por esta razón, el desarrollo de la supercomputación ha sido notable a nivel internacional, lo que se ha podido comprobar observando cómo muchos países han incrementado notablemente sus inversiones en la adquisición de infraestructuras computacionales de alto rendimiento para dar soporte a las tareas de I+D+i, la mejora de la innovación, y también la formación y mejora de la cualificación del capital humano con el objetivo de aplicar las mejoras de la investigación a la sociedad.

En una parte considerable de campos del conocimiento, la supercomputación es un elemento clave y no solo de mejora, para competir en la obtención de mejores resultados en las investigaciones. En la actualidad, la extensión del soporte de la computación de alto rendimiento a nuevos campos, en los que se pensaba que no podría ser una herramienta útil, es un hecho que está impulsando su desarrollo. Por este motivo, la formación en el uso de estas tecnologías es un proceso innovador que afecta positivamente al desarrollo de la investigación en términos globales. A lo anterior deberíamos añadir que los efectos del uso de la supercomputación son también visibles en la mejora de diversos aspectos de la sociedad ya que, mediante el uso de estas infraestructuras, se mejoran múltiples facetas de la vida de las personas, proporcionándoles también una mejora de las condiciones de vida.

En las organizaciones, tanto públicas como privadas, el efecto de la mejora de los resultados de las tareas de I+D+i se puede apreciar en aspectos como sus resultados económicos, su ventaja competitiva y su estrategia. Debido a esto, el uso de la tecnología computacional de alto rendimiento permite alinear los objetivos de las organizaciones con el fin de apoyar sus decisiones estratégicas para la obtención de los mejores resultados sobre la base de la

innovación. En definitiva, cuestiones como el desarrollo de nuevos productos o mejoras de las prestaciones y la calidad de los ya existentes, son aspectos sobre los que incide directamente la supercomputación a través de las simulaciones virtuales que ofrecen estas infraestructuras.

El objetivo general del presente trabajo es analizar, en un sentido amplio, la evolución de la supercomputación y sus principales aplicaciones para obtener una herramienta que permita medir el grado de adecuación de la formación relacionada con la supercomputación en base a los aspectos pedagógicos, de conocimiento y de contenido. Para cumplir con el objetivo propuesto, se desarrollará un cuestionario que permita analizar la percepción de los estudiantes sobre la formación relacionada con la supercomputación, sobre la base de un análisis histórico y del uso de la supercomputación para la formación y el desarrollo de tareas en diferentes áreas de conocimiento. Para alcanzar este objetivo, la investigación se desarrolló en dos fases diferenciadas:

a) Una revisión de la literatura que consiste, en primer lugar, en un análisis de la supercomputación desde un punto de vista histórico, así como sus perspectivas de futuro y el uso de redes de comunicación científica; en segundo lugar, en relación a cuestiones relacionadas con la formación en supercomputación y, en tercer lugar, mediante un estudio de la tipología de formación específica y el estudio de un caso.

b) Obtención, sobre la base de la revisión de la literatura relacionada con el modelo TPACK, de un cuestionario que permita analizar la formación relacionada con la supercomputación, con los aspectos pedagógicos, de conocimiento y de contenido con el fin de mejorar la cualificación de los estudiantes.

El trabajo se estructura como sigue: en el primer y segundo capítulo del trabajo se describen los objetivos generales y específicos que se quieren alcanzar con la realización del estudio y la metodología empleada para su consecución. En el tercer capítulo, se analiza la revisión de la literatura realizada sobre una visión histórica de la supercomputación, su evolución y las perspectivas de futuro, así como un análisis de las redes de comunicación científica y

su relación con la supercomputación. En el cuarto capítulo del trabajo se muestra una revisión de la literatura, en este caso en relación a la formación relacionada con la supercomputación, sus limitaciones y la forma de solventarlas. El quinto capítulo analiza el detalle de la formación específica y el estudio de un caso. El sexto capítulo aborda el proceso desarrollado para la elaboración y validación de un cuestionario relacionado con la formación en supercomputación. Por último, en el capítulo séptimo del trabajo se recogen las principales conclusiones del trabajo de investigación desarrollado, así como algunas recomendaciones de futuro.

## 1.- OBJETIVOS

---

El objetivo general del presente trabajo es **analizar de forma integral la supercomputación y obtener una herramienta que permita analizar la percepción de los estudiantes sobre la formación relacionada con la supercomputación.**

Para el desarrollo del objetivo general se plantean los siguientes objetivos específicos en cada uno de los temas de desarrollo:

Tema 3 - Revisión histórica y desafíos futuros en supercomputación y redes de comunicación científica.

- O1: Determinar el momento histórico considerado como el nacimiento de la Supercomputación.
- O2: Establecer cómo y cuándo comenzó el uso de la supercomputación con fines científicos.
- O3: Describir los usos de la supercomputación y sus desafíos en relación a futuros usos.
- O4: Analizar el desarrollo de las redes de comunicaciones científicas.
- O5: Determinar el soporte que proporcionan las redes de comunicación científica a la supercomputación.

Tema 4 – Supercomputación para mejorar el rendimiento en la educación superior: una revisión de la literatura.

- O1: Determinar los factores a considerar para mejorar la formación en supercomputación.
- O2: Describir cómo la formación en supercomputación puede mejorar los resultados de los investigadores.
- O3: Conocer las limitaciones de la formación relacionada con la supercomputación.

- O4: Conocer las soluciones descritas para resolver las limitaciones relacionadas con la formación relacionada con la supercomputación.

Tema 5 – Integrando clústeres de supercomputación en la educación: un estudio de caso en biotecnología.

En este tema se pretende definir un diagrama de flujo del proceso educativo relacionado con la Supercomputación, con los siguientes objetivos específicos:

- O1: Estudio de los principales campos de aplicación de la Supercomputación para los estudiantes.

- O2: Identificación del tipo de cursos computacionales relacionados con la Supercomputación en la revisión de la literatura.

- O3: Estudio relacional del campo o sector que usa la supercomputación y el tipo de capacitación computacional específica.

- O4: Estudio de caso basado en cursos relacionados con la supercomputación.

Tema 6 – Análisis del desarrollo y la validez predictiva del cuestionario de formación relacionado con la supercomputación.

- O1: Desarrollo de un cuestionario que permita analizar la percepción de los estudiantes sobre la formación relacionada con la supercomputación.

## **2.- METODOLOGÍA**

---

En el presente apartado se va a hacer una descripción general de la metodología seguida en cada uno de los apartados que se van a desarrollar en el trabajo. Posteriormente, en cada tema se van a analizar con mayor detalle las acciones seguidas en relación al desarrollo metodológico específico.

En los temas 4, 5 y 6 se ha llevado a cabo un proceso de revisión de la literatura, en la que se ha acudido a información extraída de bases de datos especializadas (IEEEExplore Digital Library, ACM Digital Library, Elsevier ScienceDirect, Scholar Google y Web of Science, etc.) en las que se ha consultado información especializada en las materias objeto de la revisión efectuada. Dentro del proceso de revisión, y teniendo en cuenta que los campos relacionados con la tecnología son heterogéneos, también se han analizado otros materiales útiles recopilados en consultas efectuadas en sitios web e informes relevantes sobre la materia.

En los estudios basados en la revisión de la literatura, se han aplicado diversos filtros para que los resultados de las búsquedas fueran de la mejor calidad, estableciendo diversos criterios de inclusión y exclusión de los trabajos que garanticen la selección de los más relevantes. Para el desarrollo del análisis se ha contado, en todo momento, con la colaboración del personal técnico, con experiencia en áreas de formación, del Centro de Supercomputación de Castilla y León (SCAYLE), que también ha colaborado en la elaboración del estudio de caso detallado en el capítulo 6.

En el desarrollo empírico del trabajo, en el que se ha obtenido la herramienta para analizar la percepción de los estudiantes sobre la formación relacionada con la supercomputación, la metodología seguida se basó en cinco fases: 1) pool de ítems, para obtener una versión inicial de posibles preguntas, 2) validación por expertos para proporcionar opiniones de alto nivel sobre el cuestionario, 3) una prueba piloto para examinar y refinar los ítems del

cuestionario propuesto por expertos, 4) una encuesta final: adaptación de la encuesta modelo TPACK, y 5) un estudio final.

### **3.- REVISIÓN HISTÓRICA Y DESAFÍOS FUTUROS EN SUPERCOMPUTACIÓN Y REDES DE COMUNICACIÓN CIENTÍFICA**

---

#### **RESUMEN**

La supercomputación implica no solo el desarrollo y provisión de infraestructuras de gran tamaño y capacidad para la comunidad científica y empresarial, sino que supone una nueva forma de gestionar las tareas de investigación, desarrollo e innovación, para lo que es necesario utilizar redes de comunicación de alta capacidad que permiten la transferencia de grandes volúmenes de información entre los centros donde se generan los datos de investigación y las infraestructuras supercomputacionales. Inicialmente el uso de supercomputadoras se produjo principalmente en el campo militar, a través de tecnología muy rudimentaria, ofreciendo pocas posibilidades de comunicación en red. A través del desarrollo tecnológico se ha ido produciendo una notable mejora de la seguridad, la privacidad y la calidad de los servicios, principalmente en lo relativo al intercambio de información, lo que ha facilitado la creación de grandes redes de comunicación científica, que a su vez han permitido la incorporación de infraestructuras para la informática de alto rendimiento en la mejora de ciencia. En este capítulo se analiza la evolución de la supercomputación y las comunicaciones científicas, por medio de una revisión crítica de su estado actual, así como identificando los principales usos actuales y la predicción de los desafíos sobre los usos futuros de este tipo de servicios avanzados.

### 3.1.- INTRODUCCIÓN

La supercomputación es actualmente uno de los tres pilares, junto con la investigación teórica y de laboratorio, sobre en el que se basa gran parte del progreso de la ciencia y la ingeniería. Durante varios años, los estudios relativos a los sistemas distribuidos y la paralelización han sido una de las líneas de investigación más desarrolladas en el área informática (Dongarra et al., 2002). Los supercomputadores permiten realizar operaciones con grandes volúmenes de datos e implementar programas de simulación en los más variados campos de la ciencia, a través de sus elementos de procesamiento. Para facilitar el acceso a estas infraestructuras computacionales de alto rendimiento y con la finalidad de promover el funcionamiento eficiente de todo el sistema de innovación (Soete et al., 2002) se han creado centros de supercomputación que están desarrollando una nueva generación de profesionales, empresas y organizaciones relacionadas con la ciencia y la tecnología (Bethel et al., 2011).

La inversión en estas infraestructuras tiene efectos positivos en el crecimiento de la productividad económica (Guellec y Van Pottelsberghe de la Potterie, 2001; Lederman y Maloney, 2003; Romer, 1990), lo que permite al mismo tiempo la mejora de la calidad, la innovación y la competitividad (De Filippo et al., 2008). Lo anterior es necesario para mejorar la creación y explotación del conocimiento científico y asegurar la calidad de la educación superior (Bernhard, 2009), lo que significa un cambio: de "objeto de investigación" hacia un cambio en las "formas de hacer investigación" (Bermeo et al., 2001).

Actualmente, la supercomputación, como parte de *e-Science*, ha transformado la forma tradicional de trabajar de la ciencia, a través de colaboraciones globales entre investigadores, el uso de grandes cantidades de datos, el uso de redes de alta velocidad y una gran capacidad de visualización que permiten un tipo de investigación que no era posible a principios del presente siglo (Atkins et al., 2010). La *National Science Foundation* (NSF) de los Estados Unidos, a través del "Informe Atkins sobre Infraestructura

Cibernética" (Atkins, 2003), el informe denominado "Hacia la Ciencia 2020" de Microsoft (Emmott y Rison, 2008) y el mundo de la ciencia en general, han concluido que ningún científico puede ser productivo o eficiente, en términos de estándares de investigación global, si no puede integrar como un factor vinculante la supercomputación en sus procesos de investigación.

En este capítulo se analiza tanto la historia de las redes de supercomputación y comunicaciones científicas, como su evolución y desafíos futuros, especialmente debido al aumento significativo del trabajo conjunto en la comunidad científica y el proceso de globalización basado en la investigación transnacional desarrollada sobre acuerdos de colaboración e intercambio de recursos y actividades conjuntas. Para el desarrollo del capítulo se contó con la colaboración de un grupo de expertos en supercomputación científica y redes de comunicaciones para aclarar conceptos con el fin de cumplir con los objetivos establecidos. En el desarrollo del capítulo no solo se realizará un análisis tecnológico, sino respecto a los diferentes usos que se han hecho a lo largo del tiempo y han influido en su progreso, así como en lo que se espera en el futuro en varios sectores. En consecuencia, el estudio se centra en el análisis de hechos históricos respecto a las redes de supercomputación y comunicación científica, a través de una revisión sistemática, que sirva de base de futuros estudios prospectivos sobre el tema. El capítulo está estructurado de la siguiente manera: la sección 2 describe métodos y objetivos sobre los que se plantearán las preguntas de la investigación; la sección 3 detalla los principales hallazgos; la sección 4 se describe detallan las principales limitaciones del estudio.

### **3.2.- METODOLOGÍA Y PREGUNTAS DE LA INVESTIGACIÓN**

Para la realización del estudio se desarrolló una extensa revisión de la literatura, con una perspectiva histórica, de las bases de datos digitales especializadas (Scopus, Web of Science y Science Direct) para detectar, obtener y consultar artículos especializados en relación con el tema objeto de estudio. También se analizaron otros materiales útiles, como

los sitios web<sup>1</sup> e informes relevantes sobre la materia, con el fin de extraer y recopilar la información relevante. El número de referencias de la palabra supercomputación fue de 1.627 en la búsqueda realizada, de las cuales el 10% se han utilizado para realizar el estudio del presente capítulo. Los cinco objetivos propuestos se han establecido con el fin de determinar la relevancia de los aspectos de la supercomputación y la relación que la evolución de las redes de comunicación científicas ha tenido a este respecto.

Las preguntas a contestar para cumplir con cada uno de los objetivos planteados son las siguientes:

- O1: Determinar el momento histórico considerado como el nacimiento de la supercomputación. Para lo que es necesario responder a la siguiente pregunta: ¿Cuándo nació la supercomputación?
- O2: Establecer cómo y cuándo comenzó el uso de la supercomputación con fines científicos. La pregunta planteada para dar cumplimiento a este objetivo es: ¿Cómo y cuándo se hizo la transición de los usos iniciales de la supercomputación a usos científicos?
- O3: Comprender los usos de la supercomputación y sus desafíos en relación a futuros usos. La pregunta planteada para cumplir con este objetivo es: ¿Cuáles son los usos actuales de la supercomputación y cuál es el pronóstico para el futuro?
- O4: Analizar el desarrollo de redes de comunicaciones científicas. Para lo que se plantea la siguiente pregunta: ¿Cómo se han desarrollado las redes de comunicación científica?
- O5: Determinar el soporte que proporcionan las redes de comunicación científica a la supercomputación. La pregunta planteada para cumplir con este objetivo es: ¿Las redes de comunicaciones científicas ayudan a la supercomputación?

---

<sup>1</sup> [www.top500.org](http://www.top500.org) y [www.green500.org](http://www.green500.org) entre otros

### **3.3.- PRINCIPALES HALLAZGOS DE LA INVESTIGACIÓN**

En esta sección se presentan los resultados de la revisión de la literatura llevada a cabo. Las siguientes subsecciones presentarán los resultados obtenidos en relación a las preguntas planteadas para dar cumplimiento a los cinco objetivos de la investigación introducidos previamente.

#### **3.3.1.- P1: ¿Cuándo nació la supercomputación?**

En el desarrollo de la supercomputación se distinguen dos etapas: la secuencial, que comienza en la década de 1940 y la paralela, a partir de la década de 1960, continuando hasta la actualidad. Cada una de las etapas descritas se compone de 3 fases distintas: una fase de arquitectura, en relación con el hardware del sistema y dos fases de software, una relacionada con compiladores y la tercera con bibliotecas y/o paquetes de aplicaciones que permiten a los usuarios evitar la escritura de ciertas partes del código (Bacon, 1998; Banerjee, 1994; Cosnard y Trystran, 1995).

La historia de las supercomputadoras se remonta al año 1943, cuando se introdujo el “Coloso”, considerada como la primera supercomputadora en la historia, diseñada por un grupo pionero en la teoría de la computación (Hermes, 1969), cuyo objetivo era descifrar las comunicaciones durante la Segunda Guerra Mundial (Singh, 2000). En el mismo año se creó en Estados Unidos el ENIAC (Mc. Cartney, 1999), que era un integrador electrónico numérico que llegó a ser considerado como uno de las supercomputadoras más grandes de la época, con fines generales a gran escala. En 1946, la Universidad de Cambridge construyó la calculadora automática de almacenamiento electrónico de retardo (EDSAC), considerada la primera computadora programable para uso general y el primer computador con programa digital, electrónico y con amplia capacidad de almacenamiento en el mundo (Wilkes y Renwick, 1950). La arquitectura utilizada en ese período se conoce como "arquitectura Von Neumann" (Von Neumann, 1945) que todavía está en uso y consiste en un procesador capaz de leer y escribir en una memoria que almacena una serie de comandos o instrucciones y realiza cálculos basados en grandes cantidades de datos de entrada.

En las siguientes décadas, el desarrollo continuó a un ritmo muy rápido, mayor en los Estados Unidos que en Europa, como se indica en el informe de 1956 del Departamento de Investigación Científica e Industrial del Reino Unido sobre informática de alto rendimiento. En los años 50, se crearon nuevas supercomputadoras, como SEAC, ERA 1101 y ERA 1103. Más tarde, IBM también desarrolló varios modelos, siendo una de las empresas más destacadas en la creación de gran parte de las infraestructuras de esta década (National Academy of Sciences, 2005). En 1959, ocurrió un hito significativo cuando la Universidad de Manchester y la empresa Ferranti cooperaron para crear la supercomputadora conocida como Atlas (Lavington, 1978). En 1960, la primera supercomputadora comercializable llamada CDC 6600 fue lanzada al mercado (Thornton, 1970), superando con creces las computadoras más potentes de la época en potencia informática y coste. En 1962 se introdujo un supercomputador que fue 80 veces más poderoso que versiones anteriores como el Meg / Mercury y 2400 veces más poderoso que Mark 1, la otra gran infraestructura computacional de la época. A finales de los años 60, se lanzó el CDC 7600 (Saunders y Guest, 1982), que ha sido considerado por muchos como la primera supercomputadora en sentido estricto, según los estándares actuales.

La introducción de la supercomputación en la industria comenzó en los años 60, cuando se fabricaron las primeras computadoras paralelas. La mayoría de estas máquinas eran procesadores mono-vectoriales (Gengler et al., 1996). Todas las máquinas incluían memoria de circuito integrado cuyo coste era muy alto y el número de procesadores no superó los 16 (Gengler et al., 1996).

En los años 80, las supercomputadoras atrajeron cada vez más la atención científica (Fernbach, 1984), principalmente debido al fenómeno surgido con el comienzo de la computación distribuida, que produjo un aumento de 16 veces en la velocidad y la memoria principal en capacidad de equipos existentes en esas fechas. Un ejemplo de los avances de esta época fue el de la supercomputadora CRAY-2 (Simmons y Wasserman, 1990), lanzada al mercado en 1985. Esta infraestructura fue entre 6 y 12 veces más rápida que su

predecesor. En los años 90, las computadoras de alto rendimiento estaban más relacionadas con las innovaciones arquitectónicas y el software (Hwang, 1993; Stone, 1993), permitiendo resolver los problemas de paralelizar tareas complejas como procedimientos estadísticos o el uso de imágenes digitalizadas mediante el uso de nuevos algoritmos. Estos avances se han basado en arquitecturas de anillos múltiples con topología escalable, lo que permitió su uso como bloques de construcción paralelos para el empleo de algoritmos más complejos (Arabnia, 1995; Bhandarkar y Arabnia, 1995a; Arabnia y Smith, 1993).

En 1991, el Congreso de los Estados Unidos publicó la Ley HPCA de computación de alto rendimiento (High Performace Computing Act, 1991) que permitió el desarrollo de la Infraestructura Nacional de Información. En 1998 se creó la primera supercomputadora que superó la barrera de gigaflop de  $10^9$  operaciones por segundo según la prueba de Linpack (Dongarra et al., 1979). En ese mismo año, en respuesta a un informe del Comité Asesor de Tecnología de la Información de la Presidencia de los Estados Unidos (Kennedy y Joy, 1998), la NSF (National Science Foundation / USA) desarrolló varias iniciativas "TeraScale" para la adquisición de computadoras capaces de realizar miles de millones de operaciones por segundo (teraflops), discos de almacenamiento con capacidades de miles de millones de bytes (terabytes) y redes con anchos de banda de miles de millones de bits (gigabits) por segundo. Sobre la base de esta iniciativa, el proyecto TeraGrid se inició en 2001 (Catlett et al., 2006) y en 2004 entró en modo de producción completa, proporcionando servicios coordinados e integrales de soporte general a la investigación académica en los Estados Unidos. En 2005, la NSF extendió su apoyo a TeraGrid al proporcionar una inversión de 150 millones de dólares para operaciones de soporte a usuario y mejora de las instalaciones durante los próximos cinco años.

En 2006 se creó la "Task Force HPC Europe", que era un grupo de trabajo de expertos para analizar la evolución de la supercomputación en Europa. Como resultado del trabajo realizado por este grupo, se publicó un Libro Blanco para la mejora de la computación con fines científicos en Europa (Kupczyk y Meyer, 2010). Este informe fue un impulso para la

asociación PRACE (*Partnership for Advanced Computing in Europe*), lo que permitió concluir que sólo a través de un esfuerzo conjunto y coordinado, Europa será capaz de ser competitiva, principalmente debido a que los costes de los sistemas de supercomputación serán de tal magnitud que ningún país europeo de forma individual podría competir con los EE. UU. y otros países de Asia o América Latina. En el mismo sentido, el informe de IDC (2004) proporciona una serie de recomendaciones para que Europa lidere la investigación científica y la industria en 2020. En 2012, la Comisión Europea (CE) anunció un plan que incluía duplicar sus inversiones en supercomputación de 630 millones a 1.200 millones de euros (Villarubia, 2012), con la finalidad de favorecer el desarrollo de supercomputadoras 'exa-scale', para 2020, con capacidad para realizar  $10^{18}$  operaciones por segundo.

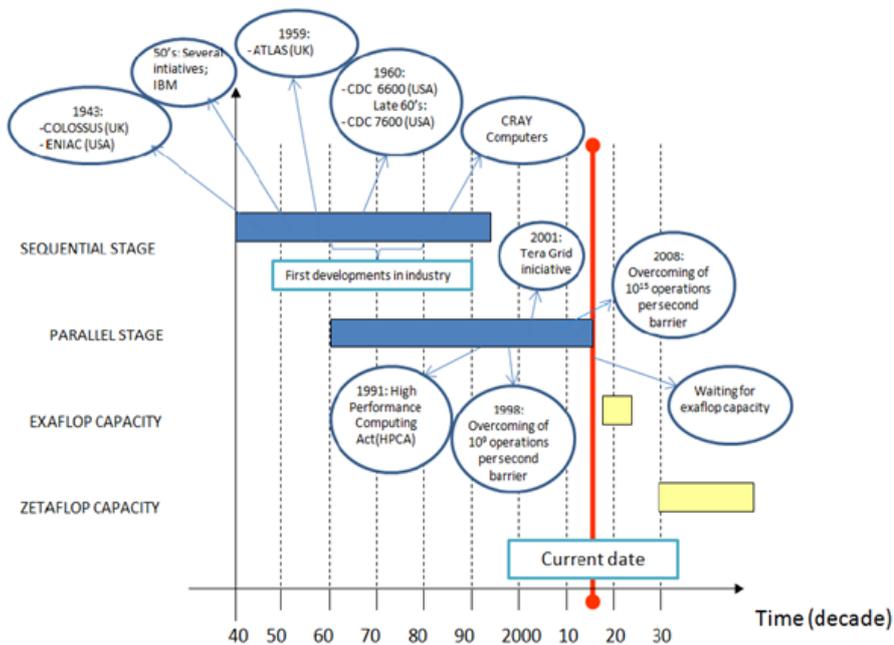
En 2008 se creó la primera supercomputadora que alcanzó la velocidad de petaflop (1.015 operaciones por segundo), cuya velocidad fue más de un millón de veces que la versión anterior (Gaudiani, 2012). Este sistema tenía casi 20.000 veces el número de procesadores de la supercomputadora más rápida existentes de hacía dos décadas y cada uno de los procesadores eran casi 50 veces más rápidos. En la primera década del siglo XXI, el escenario de crecimiento exponencial continuó experimentando un aumento de la capacidad sin interrupción, debido a factores como el efecto de Ley de Moore (Schaller, 1997), que establece que cada 24 meses las capacidades de la tecnología, tanto en la integración de transistores como en el consumo de energía se duplicarán (Balladini et al., 2010). Motivado por esta última relación, para el funcionamiento de estas infraestructuras, se requieren grandes sistemas de refrigeración, lo que es un factor limitante para la supercomputación. En los últimos años ha surgido una auténtica preocupación por la eficiencia energética. Este hecho se refleja en la creación de una lista conocida como Green 500 lista<sup>2</sup> en noviembre de 2007, que permite observar la velocidad de cálculo con menor consumo de energía de las 500 supercomputadoras más eficientes de todo el mundo.

---

<sup>2</sup> Contenido de la lista en junio de 2020: <https://www.top500.org/lists/top500/list/2020/06/>

En la Figura 3.1 se puede observar claramente los hitos principales desde el nacimiento de la supercomputación analizados a través de la revisión de la literatura, que sirven de base para predecir los desarrollos futuros en esta materia. Las líneas horizontales muestran las grandes etapas de desarrollo, las de color azul para el pasado, y las de color amarillo para el futuro. Figura 1 muestra que el uso de la supercomputación ha estado creciendo de forma continua desde 1940, ayudando al desarrollo de industria y ciencia.

**Figura 3.1: Evolución de la Supercomputación.**



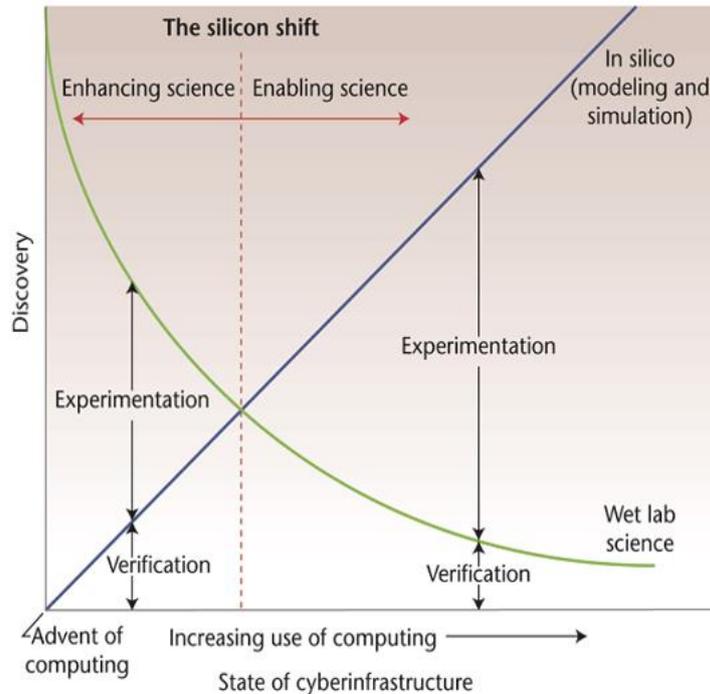
Fuente: elaboración propia

### 3.3.2.- P2: ¿Cómo y cuándo se hizo la transición de los usos iniciales de la supercomputación a usos científicos?

Antes del advenimiento de la supercomputación, la experimentación se realizaba básicamente en un laboratorio o en el campo, y la tecnología de la información y la comunicación (TIC) solo se utilizaba para ayudar en la verificación de los procesos. Como se puede observar en la Figura 3.2, con el tiempo se fue incrementando el uso de grandes infraestructuras computacionales, lo que hizo posible crear mejores modelos para llevar a cabo la simulación científica, permitiendo dedicar más tiempo para la experimentación y menos en verificación. A partir de este momento, conocido como "*Silicon Shift*" (SS),

comenzó un período en el que las grandes computadoras sirvieron no solo para mejorar la ciencia sino también para permitir el desarrollo de la ciencia.

**Figura 3.2: El momento decisivo en la ciencia: El cambio de silicio (SS)**



Fuente: Elmagarmid et al., 2008

En los años 50 y 60 hubo una gran competencia a nivel internacional entre los dos bloques, liderados por la Unión Soviética y los Estados Unidos, con el fin de liderar el mundo de forma global. En 1957, la Unión Soviética lanzó el Programa Sputnik (Hauben, 2010), que consistió en una serie de misiones espaciales no tripuladas que pretendían demostrar la viabilidad del funcionamiento de los satélites artificiales dentro de la órbita de la tierra. En respuesta, los Estados Unidos crearon la Agencia de Proyectos de Investigación (RPA), cuyo objetivo era ir más allá de las aplicaciones militares. En esos años también se creó la *National Science Foundation* (NSF) cuya finalidad era impulsar la investigación básica y la educación en todos los campos no médicos de la ciencia y la ingeniería. Este fue el momento en el que se considera que se dio el impulso real a la supercomputación (Aspray y Williams, 1994), que empezó a dejar de servir a un propósito puramente militar y se convirtió en una herramienta de apoyo a instituciones de investigación principalmente del

tipo público (universidades y agencias gubernamentales). Los primeros usuarios privados de las supercomputadoras fueron grandes empresas como las petroleras y los bancos.

En los años 60, Estados Unidos lanzó el Proyecto Apolo (Milone et al., 2002), que supuso el primer proyecto en el que se llevó a cabo un amplio uso de la supercomputación a gran escala. El objetivo del Proyecto Apolo era simular un vuelo tripulado a la luna con la finalidad de localizar un área adecuada para un posible aterrizaje en la superficie lunar. Este proyecto supuso un cambio radical en la forma en que se entendía la supercomputación debido a la gran complejidad de las simulaciones realizadas y los procedimientos operativos utilizados en la misión. A partir de este proyecto, se extendió el uso de la simulación computacional en sistemas grandes y complejos, tales como los usados en modelos de sistemas biológicos, inteligencia artificial, física de partículas, pronóstico meteorológico o el diseño aerodinámico.

En los años 80, se desarrollaron nuevos algoritmos para imágenes digitalizadas (Arabnia y Oliver 1987a; Arabnia y Oliver 1989), diseñados para el trabajo en una red de computadoras con una topología simple.

En los años 90, los sistemas informáticos distribuidos se usaban de forma más habitual para resolver problemas complejos, destacando las mejoras en la computación evolutiva, es decir, la inteligencia computacional a través de métodos que emulan la evolución natural de los seres vivos para resolver problemas de optimización, búsqueda y aprendizaje (Goldberg, 1989). En estos años, el desarrollo de algoritmos más complejos continuó mejorando las operaciones paralelas, destacando los casos de visión por computadora y el procesamiento de imágenes. El uso de redes llamadas *Reconfigurable Multi-Ring System* (RMRS) también fue desarrollado en este período. En estas redes, cada nodo tiene un grado fijo de conectividad y se muestra una arquitectura viable para el procesamiento de imágenes y resuelve los problemas de visión por computadora a través de computación paralela (Arabnia, 1990; Arabnia y Bhandarkar, 1996; Arabnia y Oliver, 1987b; Bhandarkar y Arabnia, 1995b; Bhandarkar et al., 1995). La supercomputación también se convirtió en

una herramienta indispensable para la industria a fines del siglo XX y principios del siglo XXI, así en el estudio de 2004 de *International Data Corporation (IDC)* (IDC, 2004) para explorar el uso y el impacto de los recursos de supercomputación en la industria y otros sectores, se observó que la mayoría de los encuestados indicaron que su uso era esencial para su negocio.

En la primera década del siglo XXI, los nuevos algoritmos mejoran el uso de la computación paralela. En especial se ha observado esta mejora en los algoritmos basados en la detección de bordes en imágenes tridimensionales, que están dirigidos a una red MultiRing (Arif y Arabnia, 2003).

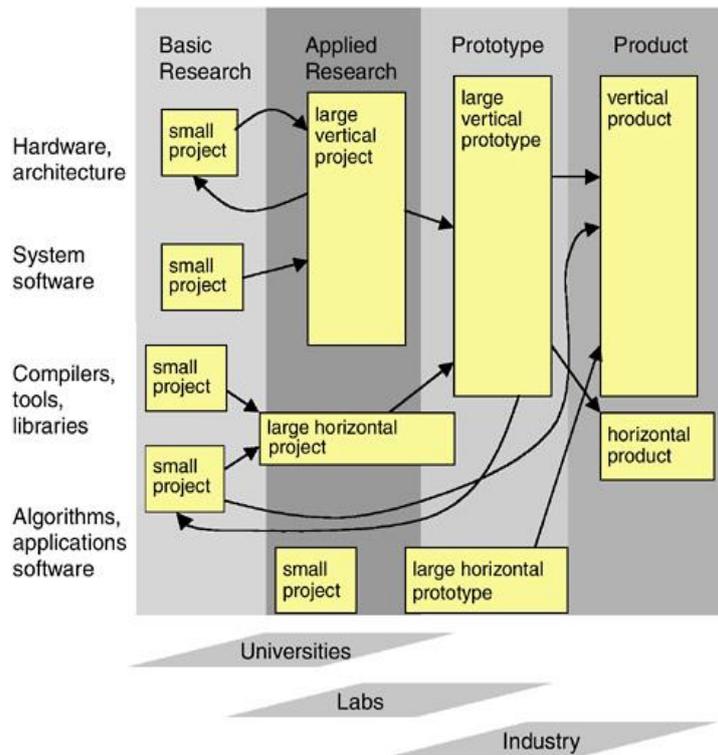
En 2007, fue posible transmitir datos cien veces más rápido usando diez veces menos de energía que para las tecnologías existentes en ese momento, por medio del uso de pulsos de luz en silicio. Esto permitió un cambio sustancial en la contribución de las supercomputadoras a la ciencia a través de la simulación y el cálculo numérico (Moraleda, 2007), como clave para ejecutar experimentos.

En respuesta a la pregunta P2, se puede concluir que los años 60 fueron el punto de inflexión para el uso de la supercomputación en el campo científico.

### **3.3.3.- P3: ¿Cuáles son los usos actuales de la supercomputación y cuál es el pronóstico para el futuro?**

La supercomputación ha revolucionado los procesos industriales, permitiendo la fabricación de mejores productos en base a un mejor análisis sobre decisiones de diseño. El uso de estas infraestructuras proporciona una reducción de tiempo y coste, no solo para un adecuado diseño del producto, sino también en el proceso de producción (Sawyer y Parsons, 2011), a través de simulaciones computacionales sobre cómo sería el producto final, lo que disminuye la necesidad de hacer prototipos. Este hecho se presenta gráficamente en Figura 3.3.

**Figura 3.3: Proceso continuo de la investigación a la producción**



Fuente: Patterson et al., 2005

En el proceso de investigación que se representa en la Figura 3.3, se pueden apreciar las etapas que van desde la investigación básica hasta la creación del producto final, realizado por los principales actores involucrados (universidades, laboratorios e industria), utilizando las herramientas de supercomputación adecuadas (hardware, software, compiladores y algoritmos). En general, la investigación básica se basa en pequeños proyectos, explorando una multitud de ideas. Los proyectos de investigación aplicada normalmente validan ideas de investigación básica y a menudo involucran a grupos de mayor tamaño. Cuando es posible desarrollar un prototipo a través de un supercomputador, que puede convertirse en un producto final, la integración de múltiples tecnologías se hace más necesaria, validando así el diseño mediante la interacción de estas tecnologías. Este desarrollo incluye muchas interacciones, mediante el cual los proyectos se inspiran entre sí, pudiendo transferir más rápidamente los resultados de la investigación básica a los productos finales, a través de múltiples iteraciones de investigación aplicada. En este contexto, los fallos se consideran

tan importantes como los éxitos para motivar tanto nuevas investigaciones básicas, como la búsqueda de nuevos productos.

La mejora de los recursos de supercomputación proporciona nuevas capacidades para la gestión y el análisis de información, así como mejores infraestructuras para archivar, conservar y explotar muchos tipos de datos, a través de los cuales los investigadores interpretan fenómenos de diferentes campos científicos. Las supercomputadoras del futuro proporcionarán un nuevo paradigma para que los desarrolladores de aplicaciones aborden nuevos desafíos a través del uso tanto de lenguajes abiertos como de otras herramientas en diversos sectores científicos y económicos (Davis et al., 2012).

La mejora de los dispositivos de tratamiento de datos, la disponibilidad de redes para distribución y la mayor capacidad de almacenamiento, ha hecho posible que las supercomputadoras puedan gestionar grandes cantidades de datos, en el rango de terabytes (un billón de caracteres) o petabytes (un cuatrillón de caracteres), e incluso mayor (exa-, zetta-, yotta- etc.). Este hecho ha sido destacado en diversas publicaciones científicas, como, por ejemplo, en una edición especial de Nature en 2008 bajo el título "Big Data: Bienvenido al petacentro, la ciencia en la era de los petabytes". En los últimos años también se destacan una gran cantidad de iniciativas internacionales, que se basaron en la anticipación de hardware exa-scale (Attig et al., 2011; Geller, 2011; Munetomo, 2011).

Los desafíos más importantes de la ciencia (Bollen et al., 2011) y la ingeniería, tanto en la simulación como en el análisis de datos, van más allá de la capacidad de los petaflops y se están acercando rápidamente a las necesidades de computación exaflop (Dongarra y Van Der Steen, 2012). El procesamiento de estos grandes volúmenes de datos plantea problemas debido a requisitos de procesamiento, que están más allá del alcance de una sola máquina (Nesmachnow, 2014), lo que marca la necesidad de mejorar el diseño de las computadoras de alto rendimiento a través de un proceso que permita el uso de toda la infraestructura de supercomputación para el desarrollo de los trabajos y no solo un determinado número de clústeres.

En la siguiente tabla (3.1) se detallan algunos usos de la supercomputación en diversas industrias y sectores:

**Tabla 3.1: Detalle de industrias y sectores que utilizan la supercomputación**

Industria del automóvil	<ul style="list-style-type: none"> <li>- Optimización de la aerodinámica.</li> <li>- Simulación del proceso de combustión de los motores.</li> <li>- Mejora de la estructura integral de los vehículos en caso de colisión.</li> <li>- Desarrollo de nuevas baterías y motores para la fabricación de vehículos eléctricos.</li> </ul>
Industria aérea	<ul style="list-style-type: none"> <li>- Simulación sobre el uso eficiente del combustible de los aviones.</li> <li>- Simulaciones, en tiempo real, de vuelo de los aviones.</li> <li>- Desarrollo de la exactitud en los modelos sobre turbulencias. En el futuro, los vuelos podrán realizarse en las condiciones meteorológicas más extremas.</li> </ul>
Industria aeroespacial	<ul style="list-style-type: none"> <li>- Simulación de la re-entrada de las aeronaves espaciales en la atmósfera terrestre.</li> <li>- Análisis del flujo de los gases en los motores de los cohetes.</li> <li>- Procesamiento de datos recibidos de satélites a gran distancia.</li> </ul>
Industria petrolífera	<ul style="list-style-type: none"> <li>- Simulaciones para encontrar reservas de petróleo y gas, a través de análisis sísmico de las olas.</li> <li>- Mejora de la exactitud para la determinación de estructuras geológicas que mejoren la localización de reservas de petróleo.</li> </ul>
Industria eléctrica	<ul style="list-style-type: none"> <li>- Manejo en tiempo real de redes eléctricas.</li> <li>- Planificación para la gestión de los activos de las redes de distribución eléctricas.</li> <li>- Medición de las redes eléctricas y su situación, así como la monitorización de tareas e instrumentos.</li> </ul>
Industria cinematográfica	<ul style="list-style-type: none"> <li>- Progresos en el campo de imagen generada por computadora (CGI).</li> <li>- Mejoras en los gráficos para el desarrollo de videojuegos.</li> <li>- Reproducción (renderizado) de imágenes para reducir el tiempo de realización de películas de dibujos animados.</li> </ul>
Sector financiero	<ul style="list-style-type: none"> <li>- Estimación del valor de los activos para la gestión del riesgo.</li> <li>- Obtención, en tiempo real de datos para mejorar la seguridad y fiabilidad de las inversiones realizadas.</li> <li>- Aplicaciones para el desarrollo, a través de la integración con terceras partes, incluidas bases de datos.</li> </ul>

Sector de la salud	<ul style="list-style-type: none"> <li>- Mejora de las simulaciones que ayuden en el diagnóstico y tratamiento de enfermedades.</li> <li>- Optimización del tiempo de secuenciación del genoma, con una importante reducción del tiempo (de 13 años de hace una década a 1 día en la actualidad) y coste (de 3 billones en el pasado a aproximadamente 1,000 dólares en la actualidad).</li> <li>- Creación de herramientas para almacenamiento y análisis de datos para medicina basada en el genoma.</li> <li>- Desarrollo de bio-simuladores, para la simulación de órganos humanos con la finalidad de encontrar la cura a enfermedades.</li> </ul>
Sector del Transporte	<ul style="list-style-type: none"> <li>- Simulación de los flujos de tráfico en tiempo real para reducir las congestiones de tráfico, los tiempos de viaje y el consumo de combustible.</li> <li>- Mejora del diseño de las redes de carreteras, utilizando herramientas de inteligencia.</li> <li>- Desarrollo de infraestructuras de soporte al vehículo eléctrico, realizando simulaciones, contribuyendo al diseño de los Sistemas eléctricos para suministro a los vehículos.</li> </ul>
Construcción	<ul style="list-style-type: none"> <li>- Desarrollo de proyectos que permiten agregación de varios sistemas (por ejemplo, detección de humo, CCTV, sistemas de aire acondicionado, etc.), mientras se combinan con resultados de simulación (por ejemplo, para predecir cómo se propagaría el fuego utilizando la dinámica de fluidos computacional, o simular cómo deberían las personas evacuar un edificio en caso de catástrofe) para producir y evaluar escenarios en casos de incendios que permitan reducir el riesgo a personas y propiedades.</li> </ul>
Emergencias	<ul style="list-style-type: none"> <li>- Análisis para soporte a la toma de decisiones que permitan anticipar situaciones que se pueden producir en un rescate mediante ensayos en tiempo real que permitan crear planes de acción que necesiten una rápida implementación. Ejemplos: accidentes de tráfico, inundaciones o incendios.</li> <li>- Creación de escenarios virtuales para simular emergencias.</li> </ul>

Fuente: elaboración propia, adaptado de Sawyer y Parsons, 2011

En los últimos años, se han descrito nuevos usos y aplicaciones de la supercomputación de las indicadas en la tabla 3.1, que permiten dar forma a las tendencias futuras en esta disciplina. A continuación, proporcionamos una descripción de los usos de supercomputación a principios del siglo XXI en diversos campos desde un punto de vista histórico, mediante el análisis de referencias bibliográficas sobre el uso de la supercomputación en la *Web of Science* en el período 2012 a 2014. Los detalles de las diferentes áreas analizadas son los siguientes:

### Campo de la Salud:

- Desarrollo de técnicas de paralelización para el análisis del genoma concurrente múltiple, que no sólo reduce en gran medida el tiempo de cálculo, sino que también da como resultado una secuencia utilizable aumentada por genoma (Puckelwartz et al., 2014). Además, se han creado nuevas técnicas para la secuenciación del ADN, como la translocación de moléculas a través de nucleótidos biológicos y nanoporos sintéticos (Martin et al., 2014), que están utilizando, entre otros, un montaje del genoma mediante técnicas de secuenciación de próxima generación (NGS) (Menhorn y Reumann, 2013). Esto permite tratamientos personalizados contra el cáncer mediante el desarrollo de técnicas de virtualización, optimizando la utilización de los recursos y la escalabilidad de NGS (Um et al., 2013). Además, se destacan los usos de herramientas para la inferencia filogenética a gran escala con supercomputadoras de máxima probabilidad (Stamatakis et al., 2012).

- En el área de cardiología se han logrado construir modelos, mediante el uso de algoritmos complejos (Wu y Cai, 2014), que muestran la interacción tridimensional completa del flujo sanguíneo en la pared arterial. Igualmente, se ha mejorado la comprensión de la función cardíaca con respecto a la salud y las enfermedades asociadas, a través de modelos computacionales multiestado muy realistas y biofísicamente detallados. Estos modelos requieren un alto nivel de capacidad computacional y algoritmos altamente escalables para reducir los tiempos de ejecución (Neic et al., 2012).

- El Proyecto Cerebro Humano (*Human Brain Project*) desarrollará una nueva estrategia integrada por medio de una nueva plataforma de investigación que integrará todos los datos y conocimientos sobre la estructura y función del cerebro humano con la finalidad de construir modelos que sean válidos para desarrollar simulaciones. El proyecto busca promover el desarrollo de la supercomputación en el campo de las ciencias de la vida y generará cuantiosos y nuevos datos neurocientíficos como punto de referencia para modelar y desarrollar nuevas herramientas para la simulación, lo que permitirá la construcción de

laboratorios virtuales para el desarrollo de servicios básicos y estudios clínicos, la simulación sobre cómo afecta el consumo de drogas y la creación de prototipos virtuales de la función cerebral mediante dispositivos robóticos (Markram et al., 2011).

- En oncología, se han descrito sistemas de diagnóstico para el cáncer de colon basados en colonoscopias virtuales por medio del procesado por algoritmos computacionalmente intensivos sobre aspectos como la preparación intestinal, la detección asistida por computadora para el examen de cáncer de colon y la detección asistida por computadora en tiempo real, con el objetivo de mejorar la sensibilidad en la detección de pólipos en el colon. Igualmente se han descrito sistemas móviles con pantallas de alta resolución conectadas a los sistemas de colonoscopia virtuales para permitir la visualización de todas las cavidades intestinales y el diagnóstico de lesiones de colon en cualquier momento y en cualquier lugar (Yoshida, 2013).

- Un estudio de 2014 describe el uso de inteligencia computacional para analizar la próxima generación de análisis genéticos a gran escala a través de la secuenciación de datos. Esto permite identificar enfermedades genéticas y los reguladores, lo cual es importante en la identificación efectiva de biomarcadores para la detección temprana en el diagnóstico del cáncer y para una adecuada planificación de los tratamientos terapéuticos para algunos tipos de cáncer (Yang et al., 2014).

- En el área de farmacia se ha observado el desarrollo de la polifarmacia, que estudia la capacidad de los medicamentos para interactuar con múltiples factores, atacando así los problemas actuales del aumento en el coste de desarrollo de fármacos y la disminución de la productividad, incorporando aplicaciones como el cribado virtual (Ellingson et al., 2014).

- En relación con el procesamiento de datos en el sector de la atención médica, se ha realizado un amplio estudio mediante “pantallas de alta resolución”, implementando mapas autoorganizados (SOM) basados en un corpus de más de dos millones de publicaciones médicas. Los resultados de este estudio muestran que es posible transformar un gran

volumen de datos en un mapa visualmente atractivo y conceptualmente relevante para los expertos (Skupin et al., 2013). En los próximos años, se estima que habrá un gran aumento en el volumen de datos biomédicos, incluida la secuenciación de la próxima generación en historias clínicas, lo que requerirá una gran capacidad de almacenamiento y modernas metodologías de cálculo (Brown y Dinu, 2013). El uso de la supercomputación en técnicas estadísticas complejas permitirá el tratamiento de datos sobre epidemiología, supervivencia y patología a nivel mundial, lo que permitirá disponer de un repositorio que permita descubrir nuevos datos sobre genética, riesgo ambiental, biología y etiología (Reumann et al., 2012).

- Las aplicaciones de alto rendimiento computacional serán útiles para proyectos de cribado virtual a gran escala, bioinformática, biología de sistemas estructurales e investigación básica para comprender el ligando proteico (Kantardjiev, 2012).

- Será posible estimar modelos biológicamente realistas sobre el funcionamiento de las neuronas, basados en datos electrofisiológicos, que es un tema clave en la neurociencia (Lepora et al., 2012).

#### Campo aeroespacial:

- La nueva generación de radiotelescopios ofrece una visión del universo con mayor sensibilidad (Wang y Harris, 2013). Por otro lado, la última generación de interferómetros para astronomía permitirá realizar estudios de datos de líneas espectrales que generarán grandes volúmenes de petabytes (Westerlund y Harris, 2014).

- Se están desarrollando simulaciones de sobre el colapso de supernovas en el núcleo de las galaxias (Lingerfelt et al., 2014). Esto fenómeno es complejo de analizar, incluso después de los extensos estudios realizados durante muchas décadas. Están sin resolver algunos de los problemas de física nuclear, como es el caso de estudios relativos a los neutrinos en condiciones extremas, así como aspectos hidrodinámicos de astrofísica (Sumiyoshi, 2011), creando así un interesante campo de estudio para el futuro.

- La supercomputación se utiliza como herramienta fundamental para las misiones de la NASA (Biswas et al., 2012) y para fines científicos y aplicaciones de ingeniería de la propia NASA (Saini et al., 2012).

- Otra aplicación importante ha sido la relacionada con el estudio de las propiedades de la convección del núcleo en el tipo A giratorio de las estrellas y su capacidad para crear fuertes campos magnéticos. Las simulaciones tridimensionales pueden servir para proporcionar datos con respecto a la asteroseismología y el magnetismo (Featherstone, 2009), al igual que la misión Kepler de la NASA, que está actualmente recolectando datos sobre la asteroseismología de cientos de estrellas (Metcalf et al., 2012). Esto permitirá entender el funcionamiento del Sol en un contexto más amplio que en la actualidad, proporcionando información estructural comparable sobre cientos de estrellas de tipo solar. También se están llevando a cabo simulaciones de datos emergentes del flujo magneto solar (Stein et al., 2012), así como nuevos avances en asteroseismología y espectropolarimetría que están comenzando a proporcionar estimaciones de rotación diferencial y magnética sobre estructuras para estrellas de tipo G y convección central en estrellas de tipo A (Toomre et al., 2012).

- La investigación en astronomía planteará serios desafíos computacionales, especialmente en la era de datos de escala de petas, de modo que algunas tareas (por ejemplo, calcular un histograma y calcular datos mínimos / máximos) no podrán lograrse sin acceso a una instalación de supercomputación que proporcione altos niveles de precisión y cobertura a través del análisis con de GPU, que son importantes en este contexto porque proporcionan una herramienta fácil de usar para la comunidad astronómica, y permite una utilización más optimizada del hardware subyacente de la infraestructura (Hassan et al., 2011).

#### Campo de la aeronáutica:

- Se han descrito desarrollos de nuevos diseños aerodinámicos a través de la simulación basada en tres apartados: el núcleo de geometría, el análisis de flujo de dinámica de fluidos computacional (CFD) y el algoritmo de optimización (Kwon et al., 2013). En estos trabajos,

los cálculos se realizan sobre la aerodinámica de los estabilizadores verticales, así como de su contribución a la estabilidad direccional y al control de aeronaves, especialmente durante la fase preliminar de diseño (Nicolosi et al., 2013). Otra aplicación notable es el estudio del flujo de aire utilizando supercomputación (Lawson et al., 2012).

#### Campo de la meteorología:

- La ejecución de simulaciones en plataformas de supercomputación permite crear modelos climáticos para predicción meteorológica, siendo clave en las tareas de investigación en 24 instituciones académicas y servicios meteorológicos en 11 países europeos (Asif et al., 2014).

- Se han descrito diversos casos de uso de software de dinámica de fluidos computacional (CFD) para modelar parques eólicos en tierra, así como la predicción y optimización de la producción agrícola a través de la asimilación de datos meteorológicos (Avila et al., 2013).

- Uno de los casos más específicos de uso de la supercomputación en meteorología es el de las simulaciones de tormentas de polvo atmosférico (Alonso-Pérez et al., 2011). Se han hecho predicciones basadas en los datos de un experimento con láser para teledetección de capas de aerosoles en la atmósfera sobre Sofía (Bulgaria), durante un episodio de Tormentas de polvo del Sahara (Stoyanov et al., 2013).

#### Campo del medioambiente:

- Desarrollo de modelos climáticos a través de la colaboración internacional multiinstitucional en el desarrollo de modelos climáticos globales (Kinter et al., 2013).

- Modelización de emisiones químicas del transporte (MCT) para estimar antropógenos y emisiones biogénicas en España con una resolución temporal y espacial de 1 hora y 1 km<sup>2</sup> respectivamente, tomando el año 2004 como período de referencia (Guevara et al., 2013).

- Asistencia en la generación de energía limpia (Kramer, 2011).

#### Campo de la biología:

- Creación de bases de datos para el análisis de genes vegetales (Yin et al., 2013).
- Desarrollo de sistemas de supercomputación paralelos para resolver problemas biológicos a gran escala, utilizando la interacción proteína-proteína (PPI) (Matsuzaki et al., 2013).

#### Campo de las emergencias:

- Desarrollo de algoritmos relacionados con la tomografía sísmica (Lee et al., 2013).
- Investigación de los ciclones tropicales y el impacto del cambio climático a través de modelos modernos basados en el trabajo de supercomputación realizado por la NASA (Shen et al., 2013).

#### Campo del sector naval:

- Predicción de situaciones reales a través de modelos tridimensionales configurados para la armada utilizando dispositivos virtuales de simulación (Bub et al., 2014).

#### Campo de la seguridad nacional:

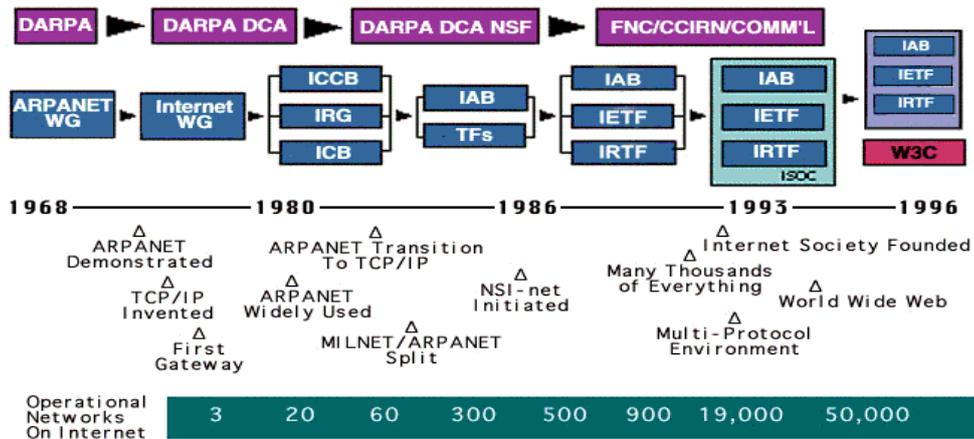
- El apoyo de las supercomputadoras será esencial en la seguridad nacional, a través del Big Data, que será utilizado para una amplia gama de casos de estudio y escenarios de aplicación, como en la lucha contra el terrorismo y el crimen, por análisis de alto rendimiento (Akhgar et al., 2015).

Las aplicaciones intensivas en datos cobrarán cada vez más importancia en el futuro. El volumen de mediciones, las observaciones y los resultados de las simulaciones aumentarán exponencialmente, de modo que el futuro esfuerzo de investigación debe centrarse en la recopilación, almacenamiento y explotación de datos, así como en el conocimiento y extracción de estas bases de datos. En resumen, en respuesta a la pregunta P3, se puede observar que prácticamente todos los campos de la ciencia y la industria experimentarán un gran avance por medio del uso de la supercomputación.

### 3.3.4.- P4: ¿Cómo se han desarrollado las redes de comunicaciones científicas?

El desarrollo de las Redes de Comunicaciones Científicas comenzó en los Estados Unidos en la década de 1960, cuando se creó la red ARPANET. Esta red de comunicación informática fue creada por el Departamento de Defensa de los Estados Unidos, cuyo primer nodo se abrió en 1969 en la Universidad de California y fue financiada por la Agencia de Proyectos de Investigación Avanzada de Defensa (DARPA), que puede considerarse la primera red de comunicación científica de la historia (O'Neill, 1995). Uno de los orígenes de estas redes radica en la carrera espacial entre los Estados Unidos y la Unión Soviética en los años 50 y 60, especialmente después del lanzamiento del satélite soviético 'Sputnik' en 1957 (Hauben, 2010).

Figura 3.4: Cronología de la evolución de ARPANET



Fuente: Leiner et al., 2009

El año 1983 se considera el del comienzo del desarrollo de Internet, tal como se conoce actualmente, en el que se produce una separación de las áreas militares y civiles de la red. En 1984 se produce un hito importante para la interconexión de supercomputadoras cuando la Fundación Nacional de Ciencias de EE. UU. (NSF) comenzó a diseñar una red que sería la sucesora de alta velocidad para ARPANET que crearía una red troncal para conectar sus seis centros de supercomputación en San Diego, Boulder, Champaign, Pittsburgh, Ithaca y Princeton. En 1986, la NSF estableció de forma permanentemente su propia red, llamada

NSFnet, que se desarrolló por los impedimentos burocráticos para usar ARPANET, que desapareció del tráfico general como tal en 1989. En ese momento, muchas instituciones ya tenían sus propias redes y la cantidad de servidores en la red superó los 100.000. Los desarrollos antes mencionados se pueden ver en la Figura 3.4.

En 1991 se aprobó la Ley de Computación de Alto Desempeño (HPCA) en los Estados Unidos, lo que permitió la financiación de una Red Nacional de Investigación y Educación (NREN). La ley se conocía popularmente como "la supercarretera de la información" y principalmente permitía el desarrollo de la informática de alto rendimiento y la comunicación avanzada, dando un impulso a muchos desarrollos tecnológicos importantes. Los expertos concluyeron que, si el desarrollo de las áreas cubiertas por la Ley se hubiera dejado a la industria privada, no habría sido posible alcanzar el desarrollo científico logrado a través de la Ley (Perine, 2000).

Desde principios de los años 90 del siglo XX en adelante, los centros de supercomputación de Illinois, Pittsburgh y San Diego contribuyeron al desarrollo de redes de alta capacidad a través de su participación en el Proyecto de Red Gigabit (Rosenberg, 1991), con el apoyo de la NSF y la DARPA. En 1995, después del final del proyecto NSFnet, estos centros se convirtieron en los primeros nodos de servicios troncales de alto rendimiento de la NFS para investigación y educación. Finalmente, el 30 de abril de 1995, NSFnet fue clausurada. Desde entonces, Internet ha consistido completamente en varios ISP (*Internet Service Provider*) comerciales y redes privadas (incluidas las redes entre universidades).

En 1996, se creó Internet 2, basado en un consorcio que surgió como una idea similar a las redes de comunicaciones científicas de los años 70, que reunió a más de 200 universidades, principalmente estadounidenses, en cooperación con 70 corporaciones líderes, 45 agencias gubernamentales, laboratorios y otras instituciones de educación superior, así como a más de 50 socios internacionales (Aulkner, 2006). Los objetivos principales del proyecto eran proporcionar a la comunidad académica una red extendida para la colaboración y la investigación entre los diferentes miembros, permitiendo así el desarrollo de aplicaciones

y protocolos que luego se pueden comercializar a través de Internet y desarrollar la próxima generación de aplicaciones telemáticas, facilitando la investigación y educación, así como la promoción de una generación de nuevas tecnologías tanto comerciales como no comerciales.

La ciberinfraestructura nacional en los Estados Unidos (Bollen et al., 2011) fue el resultado de la Ley de Investigación de Internet de Nueva Generación de 1998, la HPCA de 1991, la Iniciativa de Competitividad Estadounidense (ACI) y TeraGrid creada en 2001. En 2003, las capacidades de TeraGrid se expandieron a través de conexiones de red de alta velocidad para vincular los recursos de la Universidad de Indiana, la Universidad de Purdue, el Laboratorio Nacional de Oak Ridge y el Centro de Computación Avanzada de Texas en la Universidad de Texas, Austin. Con esta inversión, TeraGrid facilitó el acceso a grandes volúmenes de datos y otros recursos informáticos dentro del alcance de la investigación y la educación. A principios de 2006, estos recursos integrados incluían más de 102 teraflops de potencia informática y más de 15 petabytes (un billón de bytes) de almacenamiento de datos en línea y de archivos con un sistema de acceso y recuperación utilizando redes de alto rendimiento. A través de TeraGrid, los investigadores podían acceder a más de 100 bases de datos específicas para cada disciplina.

Cabe señalar que en las primeras etapas de ARPANET se hicieron pocos intentos en Europa para unirse a la nueva red, con la excepción del Laboratorio Nacional de Física (NPL), del University College de Londres en Inglaterra y el Royal Radar Establishment en Noruega (Sanz, 1998). Sin embargo, a pesar de estas iniciativas iniciales limitadas, el interés real en la tecnología desarrollada en los Estados Unidos no comenzó hasta la segunda mitad de los años 80, cuando había una gran cantidad de redes TCP / IP operando en Europa de manera aislada. Algunos de ellos comenzaron a disfrutar de las primeras conexiones transatlánticas a Internet, generalmente a través de líneas dedicadas financiadas por agencias estadounidenses como NSF, NASA y el Departamento de Energía (DoE), que estaban muy interesadas en cooperar con ciertos centros de investigación europeos. Así, en 1988 y 1989,

prestigiosas instituciones europeas en los países nórdicos (a través de NORDUnet / KTH), Francia (INRIA), Italia (CNUCE), Alemania (Universidades de Dortmund y Karlsruhe), los Países Bajos (CWI, NIKHEF) y el Reino Unido (UCL) se conectó de forma más amplia Estados Unidos con Europa. Algunas organizaciones supranacionales también establecieron enlaces dedicados a Internet en esos años, como el Laboratorio Europeo de Investigación Nuclear (CERN), la Agencia Espacial Europea (ESA) y el Grupo Europeo de Usuarios UNIX (EUUG).

Para coordinar las diversas iniciativas de redes académicas y de investigación que aparecen a nivel nacional en la mayoría de los países de Europa occidental, se racionalizaron tanto la inversión económica como las posibles soluciones técnicas. Así surgieron organizaciones como: JANET (Reino Unido), DFN (Alemania) y SUNET (Suecia) en 1984, SURFnet (Países Bajos) y AConet (Austria) en 1986, SWITCH (Suiza) en 1987, RedIRIS (España) y GARR (Italia) en 1988. Estas redes eran interdisciplinarias: su objetivo era servir a toda la comunidad académica y de investigación, independientemente de su área de actividad, mediante el uso de una única infraestructura centralizada, lo que significaba fuerzas y beneficios conjuntos de las sinergias y economías resultantes de escala.

Para optimizar el uso de estas redes, la Unión Europea está promoviendo actualmente el desarrollo tecnológico mediante el establecimiento de una red para el uso conjunto de recursos de supercomputación por parte de sus países miembros y el apoyo a estudios relacionados con la informática de alto rendimiento (Attig et al., 2011). A través de estas redes avanzadas, Europa intenta que los recursos de supercomputación sean más accesibles para los proyectos de investigación científica e industrial, y participa en importantes colaboraciones de tipo mundial que mejoran la productividad al proporcionar recursos de supercomputación para investigadores generales y científicos.

Con los años, el desarrollo de las redes de comunicaciones científicas ha ido en aumento en muchos países y continentes, además de los casos citados en los Estados Unidos y

Europa. Por ejemplo, América Latina ha estado desarrollando tales redes desde los años 90 (RedCLARA, 2011). Actualmente, hay una red llamada CLARA (Cooperación Latinoamericana de Redes Avanzadas), que apoya redes de investigación en América Latina y el Caribe, y un proyecto llamado ALICE para la interconexión entre América Latina y Europa, con el fin de crear una infraestructura para redes de investigación utilizando Protocolo de Internet (IP). Asimismo, existe una red de investigación paneuropea llamada GÉANT16, cuyo objetivo es liderar la operación a través de la asociación con cuatro NREN (Redes Nacionales de Investigación y Educación) europeas con estrechos lazos históricos y sociales con América Latina. La Figura 3.5 muestra los detalles:

**Figura 3.5: RedCLARA 2013**



Fuente: Utreras, 2014

Desde los años 90, los gobiernos de otros países, como China, ha utilizado eficazmente el potencial de investigación del sector público para impulsar la economía basada en el conocimiento (Turpin et al., 1995), financiando, de forma casi ilimitada, recursos humanos altamente cualificados, convirtiéndose en la quinta nación líder en términos de su participación en publicaciones científicas del mundo con un crecimiento exponencial en la tasa de trabajos publicados, lo que lo convierte en un actor importante en tecnologías

críticas como por ejemplo, la nanotecnología. En China se han descrito la construcción de redes de comunicación científica (Zhou y Leydesdorff, 2006) y se han llevado a cabo multitud de estudios desde la perspectiva de cómo desarrollar un sistema o entorno nacional eficaz para las innovaciones y para una mayor colaboración entre la industria y la educación superior, lo que lleva a una eficaz transferencia de conocimiento.

En Japón, el desarrollo de un nuevo sistema de investigación a lo largo de los años 90 ha llevado a la aparición de nuevos sistemas de innovación en los que se han buscado vínculos entre la universidad y la industria como un medio para estimular el crecimiento económico regional. La idea de un sistema de innovación regional (RIS) es relativamente nueva y no recibió, hasta fechas recientes, mucha atención en los marcos de las políticas. En 2004, se introdujo un cambio 'radical' (Yamamoto, 2004) en las universidades nacionales japonesas a través de la Ley de Incorporación de la Universidad Nacional (2003), lo que significó un cambio de roles para las universidades debido a la concentración de recursos en las instituciones 'de élite', y 'regionalización' de las políticas de ciencia e innovación. Esto incluyó iniciativas y políticas de "clúster" que promueven vínculos universidad-industria más amplios a nivel regional y la promoción de redes entre la industria, las universidades y los institutos públicos de investigación, mediante el apoyo a la creación de nuevas empresas y nuevas industrias (Woolgar, 2007).

### **3.3.5.- P5: ¿Las redes de comunicaciones científicas ayudan a la supercomputación?**

Durante muchos años, la gestión y el análisis de los datos producidos por las aplicaciones basadas en supercomputación fueron un componente mínimo del proceso de modelado y simulación, en el que se descuidó la gestión de los datos del usuario. Con la creciente complejidad de los sistemas, la complejidad de los datos de entrada y salida también ha aumentado. En el futuro, el volumen de datos excederá en gran medida el volumen actual, por lo que el procesamiento será muy importante, mientras se debe preservar la privacidad. Al ritmo actual de progreso, se proyecta que los sistemas de capacidad exaflop (EFLOP)

estarán disponibles alrededor de 2019 y la capacidad zetaflop (ZFLOP) en 2030 (Meuer y Gietl, 2013). Para lograr este aumento previsto, se requerirá una memoria altamente efectiva, así como el desarrollo de metodologías de programación efectivas, lenguajes y nuevos algoritmos capaces de explotar los nuevos sistemas paralelos masivos, heterogéneos con múltiples núcleos. Los patrones irregulares de comunicación no local pueden causar cuellos de botella en las supercomputadoras multinúcleo dado el mayor volumen de datos. Se están desarrollando nuevos algoritmos de paralelización eficientes, pero, aun así, el problema de la gestión del gran volumen de datos sigue siendo una de las cuestiones más complejas de la supercomputación (Winkel et al., 2012).

La NSFnet permitió una gran cantidad de conexiones, especialmente desde universidades. Aunque su objetivo inicial era compartir el uso de costosos recursos de supercomputación, las organizaciones conectadas pronto descubrieron que tenían un medio excelente para la comunicación y la colaboración entre ellas. Su éxito fue tal, que se hicieron necesarias ampliaciones sucesivas de la capacidad de NSFnet y sus líneas troncales a una tasa de multiplicación de 30 cada tres años: 56,000 bits por segundo (bps) en 1986, 1.5 millones de bps en 1989 y 45 millones de bps en 1992. En 1993 se anunció la Infraestructura Nacional de Información (NII), uno de los cuales es la Red Nacional de Investigación y Educación (NREN), una "columna vertebral" de mil millones de bps, completada en 1996. Actualmente, la Red Internet2 ofrece 8.8 Terabits de capacidad y Tecnología Ethernet de 100 gigabits en toda su superficie y conexión a una red troncal de red internacional de 100 Gbps.

Es importante señalar que la evaluación de la efectividad de las comunidades de investigación debe abordarse, no solo teniendo en cuenta los factores de producción cuantitativa y científica, sino también los factores cualitativos que influyen en la integración exitosa o no exitosa de las comunidades de investigación. Por lo general, estas plataformas están geográficamente dispersas e interconectadas por sistemas de comunicación que permiten implementar nuevas plataformas de computación en la red y

en la nube (Foster et al., 2001). Por esta razón, la programación de tareas se vuelve muy importante para gestionar diferentes usuarios y evitar largas demoras en las colas de recursos informáticos (Nesmachnow, 2014).

En 2005, el Consejo y la Comisión de la Unión Europea acordaron, mediante resoluciones, promover y alentar el crecimiento de la innovación, la investigación y el trabajo conjunto para atraer investigadores y alentar proyectos de investigación transdisciplinarios de las redes mundiales de investigación (Comunidad Europea, 2010). Actualmente, se están lanzando nuevos proyectos de investigación colaborativa que permiten una nueva forma de hacer investigación al vincular las comunidades de investigación de forma remota, a través de la ciencia electrónica o el conocimiento electrónico.

Un claro ejemplo de la necesidad de utilizar Redes de Comunicaciones Científicas conectadas a grandes capacidades informáticas es el Proyecto SKA (Matriz de Kilómetros Cuadrados), considerado como un proyecto global sin precedentes de la ciencia<sup>3</sup> en términos de su tamaño y escala en el campo de la radioastronomía, y cuya misión es construir el radiotelescopio más grande del mundo, con un área de recolección de un kilómetro cuadrado. Constituirá la mayor variedad de radiotelescopios jamás construida, así como representar un salto cualitativo en los campos de la ingeniería y la investigación. Resultará en una mayor capacidad científica, por lo que se espera que revolucione campos como la astronomía, la astrofísica, la astrobiología y la física fundamental. Los radiotelescopios se ubicarán en Sudáfrica y Australia y el procesamiento de datos por parte de las supercomputadoras se realizará principalmente en Europa y los Estados Unidos. El gran volumen de datos a manejar demuestra la necesidad de buenas redes de comunicación que permitan el transporte óptimo de datos desde el punto de recolección hasta los lugares de procesamiento, a miles de kilómetros de distancia. El proyecto se ejecutará desde enero de 2013 hasta el 31 de diciembre de 2023. Otra prueba es el proyecto del Laboratorio Europeo de Investigación Nuclear (CERN) (CERN, 2011), con un acelerador de partículas

---

<sup>3</sup> [www.skatelescope.org](http://www.skatelescope.org)

que genera grandes cantidades de información por segundo. Se ha vuelto necesario recurrir a nuevas fuentes de análisis, ubicadas en varios países.

De acuerdo con esta estrategia de desarrollo a gran escala de las comunicaciones, debe señalarse que, según Robert Vietzke, director ejecutivo de Internet 2 (Aulkner, 2006), en el futuro, Estados Unidos estará interconectado por una red que utilizará longitudes de onda de 100 Gbps. Actualmente, las organizaciones pueden trabajar con Internet2 y redes regionales avanzadas, basadas en la conexión de capa 2 de tecnología Ethernet de 100 gigabits (GE), soporte para redes definidas por software (SDN) e implementación de un modelo desarrollado por ESnet del Departamento de Energía, llamado Ciencia DMZ. Por lo tanto, se conectarán más de 200.000 centros académicos, bibliotecas, centros de salud, organizaciones gubernamentales y de investigación, lo que permitirá el transporte por red de aplicaciones especiales para la salud, la seguridad y la administración pública, y mejorará el transporte de datos para ser analizados por Supercomputadoras (Aulkner, 2006).

En respuesta a la pregunta P5, podemos concluir que el aumento en el volumen de datos a ser procesados por las supercomputadoras hoy en día, y que se espera en el futuro, requiere un desarrollo armonizado no solo de las infraestructuras computacionales, según lo estimado por algunos autores (Munetomo, 2011), sino en la capacidad de las redes de comunicación científica para el transporte de datos y su fiabilidad, así como en la consolidación de comunidades de investigación.

Este estudio se basa en una extensa revisión histórica de la literatura sobre la evolución de las redes de supercomputación y comunicaciones científicas, infraestructuras que ayudan a realizar simulaciones (Sawyer y Parsons, 2011), esenciales para el trabajo científico (Atkins, 2003; Emmott y Rison, 2008), que a su vez promoverán el desarrollo de varios sectores industriales y de multitud de campos de conocimiento.

La supercomputación será la fuerza impulsora en el desarrollo de los hitos más importantes de la ciencia. El desarrollo de la supercomputación se basa en el procesamiento de grandes

volúmenes de datos, especialmente cuando la capacidad exaflop llegue en pocos años. Será necesario implementar procesos paralelos que requieren algoritmos complejos, así como mejorar y ampliar la capacidad ofrecida por las redes de comunicación científica con una conectividad de red mejorada que permitirá que una nueva generación de aplicaciones interactúe con máquinas basadas en la computación en la nube. El gran volumen de datos utilizados en diversos campos creará nuevos desafíos y oportunidades en el modelado, la simulación y la supercomputación teórica. Los desafíos computacionales hacen posibles nuevas oportunidades para la investigación. En muchas áreas temáticas será esencial tener un especialista en cada área de conocimiento para modelar, porque si no, las instalaciones de supercomputación no serán suficientes para satisfacer los desafíos futuros para el avance de la ciencia y la tecnología.

Sin embargo, no debe olvidarse que el crecimiento exponencial en el poder de procesamiento de las supercomputadoras, que requiere constantes avances tecnológicos (Schaller, 1997) y ocurre cada pocos meses, está limitado por el considerable aumento en el consumo de energía relacionado con las nuevas infraestructuras. Hoy en día, el crecimiento de la capacidad está relacionado con la preocupación por encontrar una gama de servicios con el menor consumo de energía posible. El futuro alcance de desarrollo programado para los próximos años incluirá supercomputadoras "exa-scale" (Munetomo, 2011), capaces de procesar un volumen de información muy superior a los límites actuales, así como las Redes de Comunicaciones Científicas necesarias que permiten el transporte de grandes volúmenes de datos.

En resumen, el análisis de las cinco preguntas de investigación demuestra cómo, anticipando la evolución futura de las infraestructuras científicas, es posible mejorar su uso actual y obtener una visión amplia sobre el uso futuro. Además, está claro que, con el conocimiento de los diferentes usos y las posibilidades futuras, el rendimiento mejorará, no solo en las áreas de mayor uso descritas, sino también en nuevos campos aún por desarrollar.

### **3.4.- LIMITACIONES DEL ESTUDIO**

Este estudio tiene una serie de limitaciones que deben considerarse al interpretar sus resultados y conclusiones:

- Solo se han examinado las publicaciones académicas que se relacionan con las redes de supercomputación y comunicaciones científicas en revistas indexadas, así como aquellas que son presentaciones de seminarios y conferencias relacionadas con asuntos técnicos. En cualquier caso, en este y otros asuntos relacionados con las tecnologías, existe una amplia gama de información relevante, aunque informal, donde las experiencias y proyectos se detallan en blogs e informes técnicos que también podrían proporcionar información muy importante y que podrían utilizarse como un suplemento a la base del estudio.

- Algunas cuestiones relevantes que podrían ser de interés para futuras investigaciones pueden quedar sin respuesta. Para saber si las preguntas cumplen con precisión los objetivos, se solicitó la colaboración de un Grupo de Expertos en supercomputación científica y redes comunicación, con el objetivo de verificar que el enfoque adoptado fuera coherente con las respuestas necesarias para cumplir los objetivos.

- Hemos tratado de analizar el mayor número posible de estudios sobre el tema, con base en una perspectiva histórica del análisis de los principales hitos, pero ha sido imposible garantizar una inclusión del 100% de todos los estudios que podrían ser de interés, como se establece en revisiones sistemáticas. La razón de esta limitación es la gran cantidad de información existente y la carga de trabajo excesiva que esto implicaría y que, en todo caso, no siempre garantizaría referencias de calidad.

- La búsqueda se realizó principalmente a través de bases de datos digitales y las limitaciones encontradas fueron que en algunos casos las búsquedas se realizaron en función de los autores. En otros casos, solo fue posible buscar el contenido inferido de las palabras clave o el propósito del estudio, con el asesoramiento del Grupo de Expertos en Supercomputación Científica y Comunicación en Redes, cuyo perfil es más técnico que académico. El número de referencias de las citas de la palabra 'Supercomputación' fue de

1.627 en la Web of Science, de las cuales el 10% se han utilizado para realizar el estudio de este artículo.

- Debido a los numerosos campos en los que se utilizan las redes de supercomputación y comunicaciones científicas, hay una gran cantidad de estudios que analizan las infraestructuras como un medio, sin tener en cuenta el objetivo final de este trabajo de investigación o los aspectos cubiertos por los objetivos y preguntas en este capítulo. Por lo tanto, en muchos casos la información obtenida no era relevante.

## **4.- SUPERCOMPUTADORAS PARA MEJORAR EL RENDIMIENTO EN LA EDUCACIÓN SUPERIOR: UNA REVISIÓN DE LA LITERATURA**

---

### **RESUMEN**

El uso de supercomputadoras está actualmente muy extendido, constituyendo un componente esencial en muchos campos de la ciencia. El interés por la informática de alta capacidad está aumentando significativamente en una población más amplia y diversa de estudiantes de educación superior, principalmente debido a que el uso de estas infraestructuras permite a los alumnos mejorar sus habilidades y el resultado de su formación. Por esta razón, la demanda de cursos relacionados con la supercomputación aumenta continuamente. En este capítulo, proponemos, a través de una amplia revisión de estudios de tipo primario, respuestas a varias preguntas que se han considerado como una forma de conocer los contenidos más utilizados en la capacitación relacionada con la supercomputación, para lo que nos hemos centrado en los factores considerados para mejorar la formación en esta materia, con el fin, entre otros, de mejorar los resultados de los investigadores en una organización de educación superior, identificar las limitaciones de la formación relacionada con la supercomputación y proporcionar soluciones para estas limitaciones. Durante el procedimiento de búsqueda de trabajos para responder las preguntas de investigación, se consideraron 1.911 estudios en la primera selección. Mediante la definición de códigos de inclusión y exclusión, en los resultados de la búsqueda en bases de datos, se estudiaron 136 artículos publicados. Finalmente, utilizando criterios de calidad, se identificaron 34 estudios como relevantes para responder las preguntas de investigación. Se describieron varios factores, como la forma en que se organizan los cursos relacionados con la supercomputación, las adaptaciones que se están aplicando actualmente en los planes de estudio relacionados con los estudiantes de estas técnicas, el uso de la capacitación en resolución de problemas y la cualificación de los docentes, entre otros,

como los más relevantes, así como varias limitaciones de este tipo de formación y la identificación de soluciones para estas limitaciones. Los datos se recopilaron mediante la búsqueda de palabras clave en las bases de datos más importantes utilizadas en “*Computational Science*”, relacionadas con la formación y la educación basada en la supercomputación y se encontraron pruebas empíricas para respaldar el efecto positivo de las computadoras de alto rendimiento (HPC) para educadores e investigadores.

## 4.1.- INTRODUCCIÓN

El concepto de supercomputación es dinámico, podría describirse como la herramienta computacional de más alto rendimiento disponible en un momento dado, que cambia a medida que se producen avances en la informática y la tecnología (Infante, 1986). Como se ha analizado en apartados anteriores, la historia de las supercomputadoras se remonta a la década de los años 40 del siglo XX, cuando se introdujeron las primeras computadoras de alto rendimiento (HPC). En los años 60, hubo un cambio radical en la forma en que se entendieron estas infraestructuras, debido a las simulaciones altamente complejas utilizadas en muchos proyectos y su introducción en la industria, cuando se construyeron las primeras computadoras paralelas (González et al., 2015). Durante las últimas décadas, la práctica educativa se enfrenta a una creciente adopción de herramientas informáticas para apoyar el proceso de aprendizaje (Jonassen, 1999; Smeets, 2005). En esta línea, la rápida introducción de las tecnologías de supercomputación en la educación (Bote-Lorenzo, et al., 2008) es un desafío clave que ofrece una gama más amplia de oportunidades, principalmente debido a la posibilidad de gestionar un gran volumen de datos e información, que han llevado a educadores e investigadores a adoptar una visión pedagógica para promover, de la mejor manera posible, la enseñanza y el uso de estas infraestructuras. Todo esto crea tareas nuevas y críticas para el Sistema de Educación Superior, que requieren de una capacitación extensa y completa, utilizando muchas herramientas para la resolución de problemas (Sung et al., 2003), en la misma línea de aprendizaje basado en la indagación, para fomentar en los estudiantes la curiosidad y motivación (Specht, et al., 2013). Estas infraestructuras pueden ayudar a los estudiantes a desarrollar su capacidad, no solo para usuarios informáticos expertos, sino también para usuarios no expertos, contribuyendo a crear una nueva generación de profesionales, empresas y organizaciones relacionadas con la Ciencia y la Tecnología (Bethel et al., 2011), de hecho, la supercomputación se considera uno de los tres pilares, junto con la teoría y la investigación de laboratorio, en la que se basa gran parte del progreso de la ciencia y la ingeniería. Por esta razón, los laboratorios HPC están desarrollando e implementando,

desde hace mucho tiempo, estrategias para transferir herramientas computacionales y tecnologías analíticas para uso de todos los sectores de la sociedad (Bruce et. Al., 1997). Actualmente, para impulsar la innovación y apoyar la investigación científica, la inversión en capacitación para una fuerza laboral mejor capacitada y con conocimientos, capaz de utilizar la tecnología de supercomputación, es esencial (Coveney et al., 2012).

El objetivo del presente capítulo es crear una revisión basada en un mapa visual, sobre la formación en relación con la supercomputación, analizando cómo la integración de esta tecnología en la enseñanza y el aprendizaje tiene efectos positivos en la motivación, las actitudes, los logros y las interacciones entre los estudiantes (Yang et al., 2011). A través de una amplia búsqueda en la literatura, no se encontró ninguna revisión en relación con el tema, lo que demuestra la poca atención prestada a este campo, a pesar del importante papel de estas infraestructuras en el aprendizaje. Para respaldar el estudio, consideramos que la clasificación de los resultados en un campo donde el alcance no ha sido descrito, categorizado o evaluado (Hammick, 2005), ayuda a los investigadores a comprender el cuerpo de conocimiento existente, proporcionando una base teórica para un estudio empírico, justificando una nueva contribución al conocimiento acumulado (Levy y Ellis, 2006), lo que es esencial para que un campo de conocimiento sea clasificado como "científico" (Hunter et al, 1982). Con el propósito de apreciar la relevancia de esta propuesta, resaltamos la importancia de las revisiones efectivas de la literatura para el avance del conocimiento en general y para la comprensión de la amplitud de la investigación sobre un tema, la evidencia empírica, el desarrollo de teorías o la provisión de un trasfondo conceptual para una investigación posterior, identificando, finalmente, los temas o dominios de investigación que requieren un análisis más profundo (Cooper, 1988; Leedy y Ormrod, 2001). Por esta razón, es importante promover este estudio, con el objetivo de llenar ese vacío con un enfoque en cómo el uso de una supercomputadora ayuda a los estudiantes a desarrollar sus habilidades. Para el desarrollo de la investigación, se estableció una colaboración de un Grupo de Expertos en Computación Científica, con la finalidad de aclarar conceptos para cumplir con los objetivos establecidos.

Es probable que los hallazgos sean de interés para los investigadores actuales, los administradores y los líderes de la industria, en la medida que toman, con mayor frecuencia, decisiones basadas en la modelización computacional y el cálculo intensivo. De hecho, se considera que los científicos no pueden ser productivos o eficientes en términos de estándares de investigación globales si no pueden integrar la supercomputación en su proceso de investigación como un factor vinculante (Emmot y Rison, 2008). En resumen, esta revisión sintetiza la literatura relevante a través de la formación en supercomputación para proporcionar un análisis exhaustivo debido a la falta de estudios previos en este tema. Específicamente, el estudio del presente capítulo plantea las siguientes cuatro preguntas de investigación: (P1): ¿Cuáles son los factores considerados que mejoran la formación en supercomputación? (P2): ¿Cómo puede la formación en Supercomputación mejorar los resultados de los estudiantes y / o investigadores? (P3): ¿Qué se sabe actualmente sobre las limitaciones en la formación relacionada con la supercomputación? y (P4) ¿Cuáles son las soluciones descritas para resolver las limitaciones relacionadas con la formación en supercomputación?

El resto de este trabajo se estructura de la siguiente manera. En primer lugar, en la sección 4.2, describimos las preguntas de la investigación. En la sección 4.3. presentamos el método de investigación para la revisión de la literatura que describe las preguntas de investigación, los criterios utilizados para la inclusión / exclusión, las fuentes de estudios, la estrategia de búsqueda, la extracción de datos y la síntesis de los resultados. En la sección 4.4, presentamos los resultados del estudio sistemático propuesto. Finalmente, en la sección 4.5, presentamos limitaciones y discusiones de futuras líneas de investigación.

## **4.2.- PREGUNTAS DE LA INVESTIGACIÓN**

Para abordar las preguntas de investigación planteadas se requirió la recopilación de varias opiniones por parte de un Grupo de Expertos en Supercomputación formado por el personal del Departamento Técnico del Centro de Supercomputación de Castilla y León, SCAYLE (España). Su trabajo se centró en la provisión de información técnica precisa e imparcial a

través de su experiencia tanto en tareas técnicas, como en las áreas de formación y educación basadas en la supercomputación aplicada en varios campos de la ciencia y la tecnología. Los expertos discutieron simultáneamente las fortalezas y debilidades de la lista inicial de preguntas de investigación propuestas, que se clasificaron con el objetivo de hacer que la selección final se considerara adecuada para cumplir con los objetivos del estudio.

Una vez que se estableció el alcance del estudio, las preguntas de investigación se formularon con precisión para guiar la revisión de la literatura (Fink, 2010). Especificamos 4 preguntas de investigación para caracterizar el concepto de formación y educación en supercomputación:

- (P1): ¿Cuáles son los factores considerados que mejoran la formación en supercomputación?
- (P2): ¿Cómo puede la formación en supercomputación mejorar los resultados de los estudiantes y/o investigadores?
- (P3): ¿Qué se sabe actualmente sobre las limitaciones en la formación relacionada con la supercomputación?
- (P4) ¿Cuáles son las soluciones descritas para resolver las limitaciones relacionadas con la formación en supercomputación?

### **4.3.- METODOLOGÍA**

La metodología llevada a cabo para el desarrollo del presente estudio ha sido la definida en las etapas de estrategia de búsqueda y fuentes de datos, criterios de inclusión y exclusión, búsqueda y selección de resultados, y evaluación de la calidad de los artículos, que se detallan a continuación.

#### **4.3.1. Estrategia de búsqueda y fuentes de datos.**

El primer filtro aplicado en la estrategia de búsqueda de la investigación consistió en búsquedas desde los años 2005 hasta 2017. La decisión de consideración de este período

fue motivada por la constatación de que desde 2005 se realizaron multitud de estudios relevantes sobre la integración de la tecnología en la educación y el uso de recursos actualizados de las nuevas tecnologías de hardware y software (Barbara y Donna, 2005; Volman, 2005; Mustafa, 2005; Bomsdorf, 2005; Lim, 2005). Este hecho podría considerarse como una característica clave para determinar el momento a partir del cual se puede analizar la difusión de la tecnología en la educación, a la vez que se puede analizar cómo la tecnología es efectiva y eficiente para los fines de enseñanza y aprendizaje (Gülbahar, 2005). El segundo filtro fue la selección de artículos escritos en inglés (Hammick et al., 2010), al ser el idioma más común a nivel internacional en las publicaciones científicas. La estrategia se completó de la siguiente manera: (1) búsqueda de los términos de principales; (2) verificación de las palabras clave en documentos relevantes y ampliamente conocidos y (3) búsqueda de formas alternativas de los términos estudiados, como sinónimos y palabras clave relevantes. Después de eso, utilizamos los operadores booleanos OR y AND para agregarlos a la búsqueda.

La decisión final sobre los términos utilizados en la búsqueda se tomó en colaboración con expertos en el campo de la supercomputación. Los siguientes términos fueron los considerados como más adecuados para buscar las respuestas a cada pregunta: "curso de informática de alto rendimiento", "formación en informática de alto rendimiento", "calidad en la educación" + "informática de alto rendimiento", "calidad en la educación" + "supercomputación", "Formación de calidad" + "Informática de alto rendimiento", "Formación de calidad" + "Supercomputación", "Curso de supercomputación" y "Formación de supercomputación". La búsqueda de estos conceptos se realizó consultando las siguientes bases de datos, que se consideran como los mejores repositorios para búsquedas sobre ciencia computacional: *IEEEExplore Digital Library*, *ACM Digital Library*, *Elsevier ScienceDirect*, *Google Scholar* y *Web of Science*.

#### **4.3.2.- Criterios de inclusión / exclusión.**

El segundo filtro fue el criterio de inclusión y exclusión, realizado sobre una selección de los estudios primarios más importantes de la literatura.

Inicialmente, la selección fue de 1.911 documentos. Posteriormente, se excluyeron 1.783 artículos, y un tercer filtro de los trabajos consultados consistió en una selección exclusivamente de aquellos documentos que abordaban enfoques educativos y / o métodos de enseñanza, lo que redujo el conjunto de trabajos analizados a 128 manuscritos, agregando otros 8 estudios seleccionados en una búsqueda manual que se estimaron muy relevantes para el estudio, totalizando 136 manuscritos como los más relevantes para responder las preguntas de investigación (Ver Fig. 4.1).

La siguiente lista muestra los criterios de inclusión y exclusión adoptados.

##### Criterios de inclusión:

- Estudios en los que se analiza la formación relacionada con la supercomputación.
- Usos de la supercomputación en relación con la formación adquirida.
- Inclusión de encuestas y/o cuestionarios sobre la satisfacción de los estudiantes en sus cursos de supercomputación.
- Escritos en inglés.

##### Criterio de exclusión:

- Escasa relación con la supercomputación (relacionada más con otras tecnologías como la computación en la nube o algo similar).
- Documentos que son meras descripciones del contenido de un curso.
- Documentos con descripciones superficiales de cursos de supercomputación.
- Falta de referencias a entrenamiento y supercomputación en el título o en el resumen.

- Artículos cuyo propósito es el desarrollo de programas relacionados con la supercomputación.
- Estudios que se repiten en más de una base de datos.
- Estudios duplicados con los mismos resultados. En este caso considerado el más completo.

#### **4.3.3.- Fases de búsqueda y selección de estudios**

Como se puede observar en la figura 4.1, la búsqueda se divide en dos fases, que fueron las siguientes:

- Fase de búsqueda 1: implica la búsqueda individual en las bases de datos electrónicas previamente seleccionadas para constituir un conjunto de documentos candidatos a ser seleccionados como relevantes para el estudio. La identificación de artículos potencialmente adecuados se basa en:

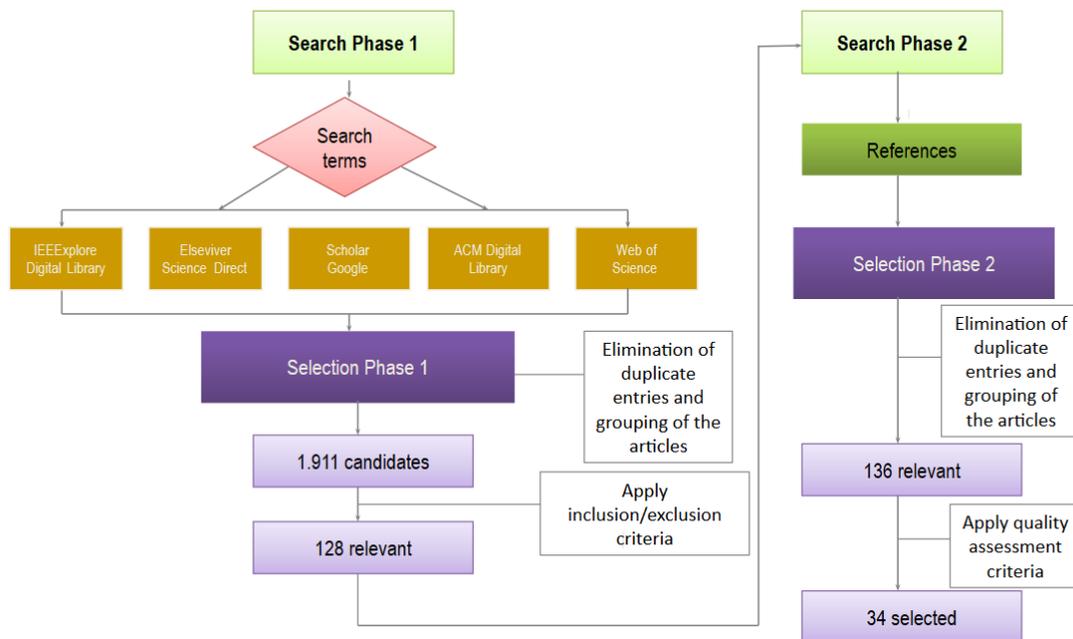
- (a) Determinación del conjunto de restricciones relacionadas con el idioma de publicación.
- (b) Definición de fuentes de estudios primarios a obtener.
- (c) Recopilación de títulos y resúmenes potencialmente útiles de los estudios primarios obtenidos.

- Fase de búsqueda 2: los artículos más relevantes se recopilan en las listas de referencias previamente seleccionadas, considerando otros documentos importantes, para el objetivo del estudio, siguiendo estos pasos:

- (a) Aplicar criterios de inclusión y exclusión a los títulos y resúmenes de los artículos seleccionados, para identificar artículos relevantes que proporcionen datos para responder las preguntas de investigación.
- (b) Extraer artículos que pudieran ser potencialmente adecuados sobre la base de títulos y resúmenes elegibles.

(c) Evaluar la calidad de la selección de documentos relevantes para obtener la selección final.

**Figura 4.1: Fases de búsqueda y selección**



Fuente: elaboración propia

#### 4.3.4.- Evaluación de la calidad del artículo.

La evaluación de la calidad se puede realizar de varias formas. La elección del método de evaluación de esta investigación se basó en una combinación de métodos, debido al alcance del estudio, tanto por su contenido técnico como educativo. Los trabajos utilizados como base para la selección final de los artículos son los especificados por Wen et al. (2012) y las categorías establecidas por Kirkpatrick (1967), revisadas posteriormente por académicos del ámbito de la educación (Harden et al. 1999; Belfield et al. 2001; Tochel et al., 2009; Hooper et al., 2013) para ayudar a evaluación de la calidad de los artículos relacionados con cuestiones educativas.

El análisis de la calidad de los trabajos evalúa la rigurosidad y la credibilidad de los estudios relevantes, en relación con su capacidad e idoneidad para responder las preguntas de investigación y su posible impacto en las conclusiones del estudio. Para obtener la selección

óptima de artículos, se han definido 10 preguntas de evaluación de calidad en la línea de los trabajos mencionados anteriormente, que son las siguientes:

P1: ¿El documento se basa en una investigación sobre formación en supercomputación o es simplemente una descripción de los usos de la supercomputación como herramienta en un campo de conocimiento?

P2: ¿Existe una declaración clara de los objetivos de la investigación, que abarca las opiniones de los participantes sobre la experiencia de aprendizaje, su organización, presentación, contenido, métodos de enseñanza, ...?

P3: ¿Se definió bien el contexto del estudio para saber cómo se llevó a cabo la investigación?

P4: ¿El diseño de la investigación fue apropiado para abordar los objetivos de la investigación, describiendo los cambios en las actitudes, creencias o percepciones como resultado de la capacitación en supercomputación?

P5: ¿Existe una descripción adecuada de la transferencia del aprendizaje al lugar de trabajo o la disposición de los alumnos a aplicar nuevos conocimientos y habilidades?

P6: ¿Hay algún cambio, debido a la capacitación en supercomputación, en la práctica organizacional?

P7: ¿Las conclusiones del estudio son claras y útiles para el objeto de la investigación?

P8: ¿Se muestran claramente las limitaciones del estudio?

P9: ¿Existe una descripción clara de la adquisición de conceptos, procedimientos y principios?

P10: ¿El estudio de valor para analizar el beneficio para clientes / usuarios es un resultado directo del programa educativo?

Cada pregunta tenía tres respuestas opcionales: “Sí”, “Parcialmente” o “No”. Estas tres respuestas se puntuaron de la siguiente manera: " Sí " = 1, " Parcialmente " = 0.5 y " No "

= 0. La puntuación de calidad de los trabajos se calcula sumando los puntajes de respuestas a las preguntas de control de calidad.

Finalmente, para garantizar que solo los ítems más relevantes tuvieran una calidad aceptable, se realizó una puntuación de cada estudio para analizar la confiabilidad de los resultados de la revisión. El requisito era recibir una calificación superior a 5.5 (nivel medio) de acuerdo a Kirkpatrick (1967). Como resultado, 34 del total de 136 artículos fueron aceptados (24,98%) y 102 (75,02%) fueron rechazados, ya que tenían un puntaje de calidad inferior a 5 en la segunda fase de selección (ver Tabla 4.1).

**Tabla 4.1: Puntuación de relevancia de la calidad de los artículos analizados**

<b>Nivel de calidad</b>	<b>Nº de estudios</b>	<b>Porcentaje</b>
Muy alto ( $8.5 \leq \text{puntuación} \leq 10$ )	3	2,20%
Alto ( $7 \leq \text{puntuación} \leq 8$ )	11	8,08%
Medio ( $5.5 \leq \text{puntuación} \leq 6.5$ )	20	14,70%
Bajo ( $3 \leq \text{puntuación} \leq 5$ )	12	8,82%
Muy bajo ( $0 \leq \text{puntuación} \leq 2.5$ )	90	66,20%
<b>TOTAL</b>	<b>136</b>	<b>100,00%</b>

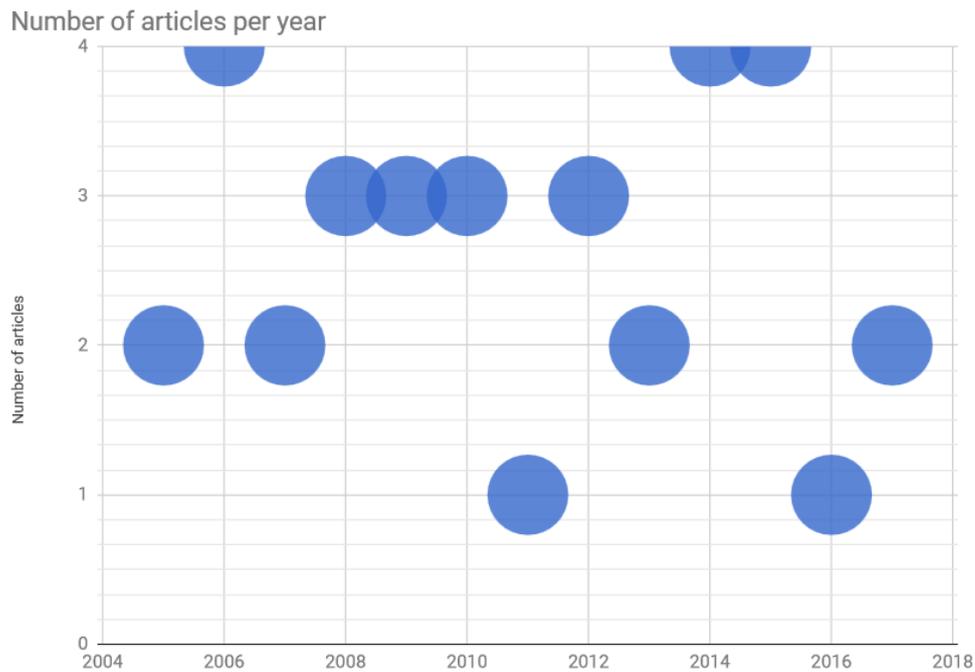
Fuente: elaboración propia

La metodología utilizada fue la de revisión sistemática específica, basada en una taxonomía de revisiones de literatura en educación (Cooper, 1988), que se considera una verdadera investigación basada en hipótesis, en la que los estudios se seleccionan y combinan con el uso de un protocolo predefinido para reducir subjetividad y la posibilidad de sesgo del investigador (Kitchenham, 2004; Anderson et al., 2011).

#### 4.4.- RESULTADOS

En el presente estudio, se identificaron 34 manuscritos, publicados durante el período comprendido entre 2005 y 2017 (ver Figura 4.2) como relevantes para responder las preguntas de la investigación.

**Figura 4.2: Número de publicaciones por año**



Fuente: elaboración propia

Las fuentes de la literatura sobre el tema objeto de investigación están ampliamente distribuidas en una variedad de publicaciones con diferentes características (como se puede observar en la Tabla 4.2). El análisis de la información mencionada muestra la naturaleza multidisciplinaria del estudio.

**Tabla 4.2: Selección de estudios por tipos de publicación**

Tipo	Nº de estudios	%
<b>Revistas</b>	<b>14</b>	<b>41,18%</b>
Age (Omaha)	1	2,94%
Biotechnology progress	1	2,94%
International Journal of Engineering Education	1	2,94%

Journal of Computing Sciences in Colleges	2	5,88%
Journal of Parallel and Distributed Computing	1	2,94%
Lecture Notes in Computer Science	1	2,94%
Procedia Computer Science	6	17,65%
Procedia-Social and Behavioral Sciences	1	2,94%
<b>Conferencias</b>	<b>15</b>	<b>44,12%</b>
ACM Transactions on Computing Education (TOCE)	1	2,94%
ACM annual conference on Innovation and technology in computer science education	1	2,94%
Computational Science–ICCS 2009	1	2,94%
Computing in Science and Engineering	1	2,94%
DoD HPCMP Users Group Conference	1	2,94%
5th Global Congress on Engineering Education	1	2,94%
IEEE Transactions on Education	2	5,88%
IEEE 26th International Parallel and Distributed Processing Symposium	1	2,94%
IEEE High Performance Extreme Computing Conference (HPEC '16)	1	2,94%
International Symposium Computer Science and Computational Technology	1	2,94%
Second International Conference on Education Technology and Training	1	2,94%
TeraGrid Conference: Extreme Digital Discovery	1	2,94%
TG08-TeraGrid Conference	1	2,94%
Workshop on Education for High-Performance Computing	1	2,94%
<b>Otros</b>	<b>5</b>	<b>14,71%</b>
Web resource	5	14,71%

Fuente: elaboración propia

Las respuestas a cada una de las preguntas de investigación planteadas en líneas anteriores se describen a continuación (las referencias bibliográficas que se indican con la letra S se detallan en el APÉNDICE I):

#### **4.4.1.- P1: ¿Cuáles son los factores considerados que mejoran la formación en Supercomputación?**

Los factores considerados para mejorar la formación en supercomputación, extraídos de la revisión de la literatura, son los siguientes:

a- El estudiante explora problemas científicos reales a través de la experiencia práctica (S2, S4, S7, S9, S10, S12, S13, S15, S19, S25, S29), con ejemplos relacionados con sistemas y programas paralelos (S5, S16, S25), aprendiendo las funcionalidades más avanzadas de las computadoras, vinculando la teoría con una solución práctica de problemas.

b- El plan de estudios de los cursos relacionados con la supercomputación se han ido adaptando, incluyendo en algunos casos, como componente central, módulos de paralelismo con un sistema de clúster de alto rendimiento fácil de usar en el plan de estudios de ciencias de la computación orientado a la enseñanza (S1,27).

c- La supercomputación fomenta la exploración de la caracterización cuantitativa del rendimiento del programa en una variedad de plataformas (S15, S18).

d- Los formadores de los cursos relacionados con la supercomputación tienen características pedagógicas positivas y competencias profesionales específicas como científicos e investigadores (S11) y están cualificados para ayudar a los estudiantes con proyectos complejos (S17).

e- No es necesario que todos los usuarios tengan habilidades en el uso de una supercomputadora. De hecho, existen herramientas como las de visualización que ayudan a los estudiantes a comprender, en profundidad, la naturaleza de problemas científicos particulares (S21) a través del aprendizaje de la computación paralela (S31), acercándose al conocimiento de la forma de resolver estos problemas (S16, S21) manipulando fácilmente grandes conjuntos de datos (S20).

f- La formación en supercomputación tiene como objetivo preparar, en un estándar de alta calidad, a las personas en un proceso continuo para obtener un nivel de conocimiento y

competencia adecuado para lidiar con los sistemas informáticos más grandes disponibles (S2).

g- La capacitación en supercomputación crea una metodología de aprendizaje modelo que involucra a los estudiantes en diferentes tipos de resolución de problemas (S34).

#### **4.4.2.- P2: ¿Cómo puede la formación en supercomputación mejorar los resultados de los estudiantes y/o investigadores?**

Actualmente, la computación de alto rendimiento (HPC) se utiliza en la educación superior para capacitar a los usuarios en la resolución de una gran variedad de problemas, convirtiéndola en una tecnología crítica en campos como la ciencia, la ingeniería, las finanzas y la investigación y el desarrollo. La supercomputación puede verse como la intersección entre las fronteras de la ciencia y la informática, y presenta a los estudiantes conceptos básicos como la simulación a gran escala.

En la revisión, identificamos algunas características de la formación en supercomputación, que están directamente relacionadas con la mejora de los resultados de las investigaciones en las organizaciones de educación superior. Son los siguientes:

a- La supercomputación permite a los usuarios obtener competencias para trabajar más profundamente en diferentes problemas, relacionados con sus campos de investigación (S4, S15) a través de experimentos reproducibles (S5), que utilizan códigos de dificultad creciente para ilustrar y experimentar, a través de los principales modelos de programación (S30).

b- El trabajo con una supercomputadora está a la vanguardia en investigación (S8) y aumenta el interés en realizar trabajos de investigación independientes (S29).

c- La disponibilidad de infraestructura de red, basada en una instalación de supercomputación (S26, S33), permite la colaboración entre grupos de investigación distribuidos geográficamente (S33) y permite la participación en proyectos de áreas multidisciplinares (S8, S19).

d- La formación en supercomputación conduce a una comprensión más profunda de los conceptos de matemáticas y ciencias (S34) que ayudan a los investigadores en el uso de modelos matemáticos (S3, S26, S34) para describir un problema en un campo de investigación específico, que permita extraer conclusiones y evaluar el desempeño sobre el problema propuesto.

e- Los cursos electrónicos basados en cómo ejecutar un proyecto relacionado con HPC son una forma importante de capacitación en supercomputación (S18), incorporando componentes experimentales como modelos para las aplicaciones (S22), simulaciones y visualizaciones (S23).

f- Implantar la capacitación en investigación, utilizando una supercomputadora, puede fortalecer el entusiasmo de los estudiantes por el descubrimiento de nuevas formas de uso de la tecnología, mejorando la creatividad y ampliando su cobertura de nuevos y mayores conocimientos (S13).

g- A medida que aumentan las capacidades computacionales, los usuarios realizan su trabajo con mayor rapidez, logrando la solución a problemas cada vez más complejos e implementando más modelos de investigación en la misma cantidad de tiempo (S20).

h- La visualización de alta calidad puede conducir a un nuevo nivel de comprensión de cómo se pueden dividir y procesar los datos en paralelo para simplificar la interpretación de los resultados de la investigación (S21). También representa un mecanismo para fortalecer la sinergia experimental-computacional (S3).

i- La simulación numérica que ofrece una supercomputadora permite el estudio de sistemas complejos y fenómenos naturales que serían demasiado costosos o incluso imposibles de estudiar por experimentación directa (S14).

j- Las simulaciones y la práctica del diseño generan grandes volúmenes de datos (S34) y preparan a los estudiantes para futuras investigaciones en el área de sistemas de bases de datos de gran tamaño (S1), que cubren una gran variedad de aplicaciones (S22).

k- La supercomputación contribuye a inventar nuevas herramientas y nuevas filosofías de trabajo, que mejoran el aprendizaje sobre el campo de la informática (S9).

l- El conocimiento en la supercomputación fomenta el diseño de programas eficientes que ofrecen a los estudiantes una comprensión crucial sobre la importancia de los algoritmos subyacentes en la implementación de programas de gran tamaño (S9).

m- La retroalimentación sobre el desempeño de los estudiantes en los cursos relacionados con la supercomputación ayuda a guiar en la capacitación, maximizando el tiempo que usan para sus investigaciones, lo cual es altamente deseable en los proyectos de investigación (S17).

#### 4.4.3.- P3: ¿Qué se sabe actualmente sobre las limitaciones en la formación relacionada con la supercomputación?

Considerando la respuesta de las preguntas de investigación de este estudio, identificamos algunas limitaciones que deben mencionarse. Para una mejor organización y comprensión de la información, las limitaciones se agruparon en cuatro tipos (ver Tabla 4.3), dependiendo de la dimensión involucrada: Estudiantes, Formadores, Infraestructuras HPC y Metodología y desarrollo de la capacitación.

**Tabla 4.3: Limitaciones en la formación relacionada con la supercomputación**

<b>Grupos</b>	<b>Limitaciones</b>	<b>Referencias</b>
Estudiantes	Insuficientes antecedentes formativos con diferentes grados o especializaciones, principalmente en lo relativo a lenguajes de programación.	S4, S9, S17, S23, S26, S28
	Restricciones de los curriculums académicos en relación a la supercomputación.	S12, S17, S25
	Falta de habilidades básicas para el diseño, implementación y operación en sistemas de supercomputación, especialmente para el desarrollo de aplicaciones de software paralelo para alcanzar mayor eficiencia, desarrollo y escalabilidad.	S1, S15, S29
	Necesidad de mayor sofisticación matemática que complemente el contenido de un curso de computación avanzada para utilizar modelos matemáticos de forma adecuada.	S32

Profesores	Se describe un substancial esfuerzo y dedicación para una exitosa implementación de las herramientas necesarias para desarrollar de forma adecuada los cursos, lo que significa menos tiempo para preparar contenidos adecuados.	S13
	Necesidad de complementar el entrenamiento de los profesores, especialmente en lo relativo a computación paralela, mediante la implicación de expertos.	S2, S11
	Falta de experiencia en el campo de la supercomputación.	S1
	Dificultad para el reclutamiento de suficientes tutores para mantener una ratio adecuado tutor/alumno.	S2
Infraestructuras de supercomputación	Rápidos cambios en los sistemas y software, tanto actualmente, como en los próximos años, con previsión de un continuo crecimiento en el volumen de datos a gestionar.	S2, S17, S18, S20, S27, S29
	Necesidad de predecir la utilidad y relevancia de las tecnologías del futuro, tanto en el corto, como en el largo plazo.	S28
	Baja disponibilidad de herramientas de supercomputación debido a los altos costes.	S8
	Escasa disponibilidad económica para la compra de equipos de computación paralela por su elevado coste.	S29, S31
	Los usuarios no son correctamente atendidos por su Centro de Supercomputación local.	S1, S28
	Dificultades para gestionar, por parte de los Centros de Supercomputación, al alto número de estudiantes con escasa experiencia en el uso de estas infraestructuras.	S2, S30
	Problemas con el uso de los interfaces de los sistemas operativos.	S13
Metodología y desarrollo de la formación	Falta de desarrollo de formación on-line para estas tecnologías.	S2
	Material para la formación inadecuado para el desarrollo de los cursos.	S2, S20, S28
	Necesidad de centrarse en mejorar las habilidades en el uso de las tecnologías de la supercomputación en prácticamente todas las áreas.	S1, S17, S28
	Restricciones de tiempo, tanto para estudiantes como para profesores, principalmente para la implementación de algoritmos numéricos.	S13, S20, S25, S28, S30
	Falta de entrenamiento en relación a programas para el uso de infraestructuras de supercomputación para utilizar de forma adecuada la programación paralela.	S24, S25
	Necesidad de emplear de forma efectiva y masivamente la paralelización en la computación científica.	S29

Fuente: elaboración propia

**4.4.4.- P4: ¿Cuáles son las soluciones descritas para resolver las limitaciones descritas en relación con la formación en Supercomputación?**

Siguiendo la misma estructura de la pregunta anterior, para una mejor organización y comprensión de la información, las soluciones de las limitaciones se agruparon en cuatro tipos (ver Tabla 4.4), dependiendo de la dimensión involucrada: Estudiantes, Formadores, Infraestructuras HPC y Metodología y desarrollo de la formación. Considerando la respuesta de las preguntas de investigación de este estudio, identificamos algunas soluciones a limitaciones que deben mencionarse en cada tipo.

**Tabla 4.4: Soluciones a las limitaciones en formación relacionada con la supercomputación**

<b>Grupos</b>	<b>Soluciones a las Limitaciones</b>	<b>Referencias</b>
Estudiantes	Implementación de una amplia gama de cursos introductorios para estudiantes con escasos antecedentes en conocimientos relacionados con la tecnología.	S9, S19, S32
	Reformas de los curriculums educativos (incluyendo materias relacionadas con la computación paralela en etapas iniciales de los estudios superiores), de los métodos de enseñanza, del tipo de evaluación y de los modelos de cursos teóricos.	S13, S14, S29
	Atracción del interés de los estudiantes a través de varias formas como los seminarios de verano, uso de robots, etc.	S10
Profesores	Incremento en el número de horas de atención a alumnos por parte de los profesores.	S19
	Ambientes de enseñanza adecuados que provean herramientas para conocer el vocabulario y los conceptos básicos relacionados con la supercomputación.	S20
	Fomento de las sesiones de formación presencial desarrolladas por expertos en cada uno de los campos, lo que tiene amplia aceptación, a la vez que mejora la transferencia del conocimiento.	S2, S28
	Ambiente de trabajo en equipo en la formación relacionada con la Supercomputación, lo que permite un mejor intercambio de ideas con la coordinación del tutor.	S7, S8, S25, S27, S31
	Transmisión de habilidades que permiten mejor desarrollo de la práctica a través de experimentos.	S5
Infraestructuras HPC	Los sistemas de supercomputación son compartidos por múltiples usuarios que interactúan de forma remota, lo que supone una solución a los altos costes de utilizar un supercomputador.	S20, S29

	Implementación de software libre y sistemas abiertos para reducir costes y permitir la facilidad de uso a través de interfaces amigables para el usuario.	S1, S6, S12, S30
	Uso de software nuevo y fácil de usar, lo que hace que las simulaciones computacionales sean accesibles a grupos más amplios de usuarios.	S34
	Uso de potencia computacional a un coste accesible, a través de alternativas a las plataformas de supercomputación, como la computación basada en clústeres, que tienen un alto grado de flexibilidad.	S1
	Uso de la tecnología <i>grid technology</i> en entornos educativos que puede ayudar a los estudiantes en el uso de modelos complejos y simulaciones.	S22, S30
	El avance de las redes tecnológicas ha permitido el uso de tecnología de alto rendimiento por parte de empresas favoreciendo el aprendizaje de computación paralela.	S11
Metodología y desarrollo de la formación	Desarrollo de materiales de estudio adecuados: tutoriales, textos en web, guías de usuarios, presentaciones electrónicas, cursos interactivos, libros y revistas que proveen conocimientos relacionados con la supercomputación de forma rápida en entornos que cambian a gran velocidad.	S2, S6, S7, S8, S9, S12, S13, S14, S17, S18, S25, S28, S30
	Uso de herramientas pedagógicas que mejoran la confianza de los estudiantes.	S5
	Provisión de acceso real a una Supercomputadora que sirve como elemento de inspiración y motivación a los estudiantes, ofreciéndoles la posibilidad de implementar algoritmos utilizando diferentes tipos de arquitecturas y modelos de programación paralela, que mejora el rendimiento de los estudiantes, permitiéndoles desarrollar estudios comparativos.	S12, S20, S30
	Colaboración de los Centros de Supercomputación en la provisión de una plataforma que permita acceder a la misma con facilidad, con el soporte de diferentes instituciones, que permita asegurar que las necesidades y expectativas son conocidas por todos los usuarios, sin interferencias en las actividades de investigación que realizan de forma ordinaria los Centros de Supercomputación.	S2, S7, S10, S14, S28, S30
	Amplia variedad de programas de formación con diferentes materias que complementan el conocimiento, de acuerdo a las habilidades que se espera obtener al final de los programas formativos.	S14, S26
	Uso del Benchmarking para mejorar la formación, a través de la descripción de una taxonomía que permita comparar diferentes tipos de experimentación para tener acceso a un rango mayor de prestaciones computacionales.	S5, S8, S14

	Desarrollo de una enseñanza orientada a los métodos de investigación, para adquirir una mayor calidad de la enseñanza e involucrar al estudiante en este tipo de aprendizaje, sobre la base de entornos de aprendizaje de investigación basados en el uso de un supercomputador.	S13, S31
--	--	----------

Fuente: elaboración propia

#### 4.5.- DISCUSIÓN

Nuestros hallazgos sugieren que incluir supercomputadoras en los procesos de aprendizaje ha brindado una amplia gama de nuevas oportunidades para los estudiantes, principalmente debido a la mejora en la calidad de la capacitación obtenida, que proporciona mejores resultados en los casos prácticos, a través simulación de situaciones de la vida real. Al mismo tiempo, el estudio identifica las necesidades de aprendizaje y diseño de estrategias para satisfacer esas necesidades. Sin embargo, como se ha podido observar en la sección anterior, hay una ausencia de referencias en la literatura sobre el grado en que el uso de una supercomputadora impacta en el resultado final de una investigación. Para descubrir cómo las características de estas infraestructuras pueden mejorar pedagógicamente el proceso de aprendizaje, se necesita un esfuerzo sistemático y más estudios empíricos, tal como se puede observar en otros análisis en relación con el uso de la tecnología en el aprendizaje (Mikropoulos y Natsis, 2011). Aunque la investigación realizada señaló la necesidad de una mejor capacitación previa de los estudiantes, esta revisión proporciona otras acciones que se implementarán, como nuevos métodos de enseñanza o la necesidad de mayor facilidad de acceso a las infraestructuras de supercomputación.

Para profundizar en la investigación, el estudio implementó varios pasos para respaldar la calidad de los resultados del análisis realizado. La interpretación de los resultados finales se puede generalizar a los objetivos, acciones y estrategias para enseñar y guiar a los alumnos en su proceso de aprendizaje, teniendo en cuenta que, sin una discusión crítica y racional sobre las oportunidades de las nuevas tecnologías en la educación, existen grandes

dificultades para progresar hacia un debate real sobre el impacto de estas infraestructuras computacionales.

#### **4.5.1.- Resultados**

Los principales resultados que responden a las preguntas planteadas al inicio de este capítulo se resumen a continuación.

- (P1) Los factores considerados para mejorar la capacitación en supercomputación muestran que el más relevante es la forma en que se organiza los cursos y las adaptaciones que se están aplicando actualmente en los planes de estudio relacionados con estas técnicas, utilizando ejemplos como modelos basados en resolución de problemas reales a través de una supercomputadora. Otro factor importante es la cualificación de los formadores, que es esencial para un mejor desempeño y, finalmente, el uso de herramientas como instalaciones de visualización que proporcionan una comprensión fácil para todos los estudiantes.
- (P2) Se observa que la posibilidad de trabajar en una red de grupos de investigación multidisciplinares mejora claramente el desempeño de los investigadores. Los contactos proporcionados por estos grupos abren la posibilidad de participar en multitud de proyectos de investigación y desarrollo, utilizando técnicas de simulación numérica, que contribuyen a resolver, en un corto período de tiempo, los procesos en los que se basa la investigación, permitiendo una mejor interpretación de resultados y la mejora de las investigaciones desarrolladas.
- (P3) Las limitaciones del estudio se centran principalmente en la formación insuficiente de los estudiantes en ciertas especialidades, motivado por muchos factores como restricciones en los planes de estudios, los rápidos cambios en las tecnologías de supercomputación, la sofisticación de los modelos matemáticos e incluso otras cuestiones como una falta general de experiencia en el campo de la supercomputación que afecta a aspectos como el reclutamiento de formadores o las limitaciones de tiempo de los cursos.

Otra limitación importante es el coste de la infraestructura de supercomputación, que causa dificultades para que los estudiantes puedan acceder a los recursos de supercomputación.

- (P4) Las soluciones propuestas para las limitaciones descritas en el estudio se centran en la mejora de la capacitación previa de los estudiantes que asisten a sesiones de formación basadas en la supercomputación mediante una formación previa que les permita tener los conocimientos básicos para poder tener un adecuado aprovechamiento de los conocimientos adquiridos, disponibilidad de material completo y actualizado, y con acceso adecuado a un clúster real de supercomputación para poder hacer prácticas reales.

#### **4.5.2.- Limitaciones**

Durante la revisión de la literatura, la mayor dificultad encontrada fue la heterogeneidad en el tipo de artículos estudiados, así como en el contenido relacionado con la tipología de la formación. Como se ve en la respuesta a las preguntas y también en el estudio de los cursos, es necesario mejorar algunos aspectos básicos para una capacitación óptima como la provisión de formación básica previa para aquellos estudiantes con antecedentes insuficientes en técnicas de supercomputación, la cualificación adecuada de los formadores y un uso más amplio de las instalaciones de supercomputación. También esperamos que el presente estudio aliente a los investigadores e instituciones en general, a contribuir en el desarrollo de acuerdos, que aborden los desafíos de usar una supercomputadora como parte del conocimiento fundamental para usos científicos, mediante el uso de una infraestructura HPC real.

## **5.- INTEGRANDO CLÚSTERES DE SUPERCOMPUTACIÓN EN LA EDUCACIÓN: UN ESTUDIO DE CASO EN BIOTECNOLOGÍA**

---

### **RESUMEN**

La integración de una supercomputadora en el proceso educativo mejora las habilidades tecnológicas de los estudiantes. El objetivo del capítulo es estudiar la interacción entre STEM (Ciencia, Tecnología, Ingeniería y Matemáticas) y materias que no son STEM para desarrollar un curso de estudio relacionado con la capacitación en Supercomputación. Proponemos un diagrama de flujo del proceso para mejorar el rendimiento de los estudiantes que asisten a cursos relacionados con la supercomputación. Como resultado final, este estudio destaca el análisis de la información obtenida mediante el uso de infraestructuras HPC en cursos implementados en la educación superior a través de un cuestionario que proporciona información útil sobre sus actitudes, creencias y evaluaciones. Los resultados nos ayudan a comprender cómo la colaboración entre instituciones mejora los resultados en el contexto educativo. La conclusión proporciona una descripción de los recursos necesarios para la mejora de la Educación en Supercomputación (SE), proponiendo futuras direcciones de investigación.

## 5.1.- INTRODUCCIÓN

Las tecnologías de alto rendimiento son un tema de estudio cada vez más desarrollado en el ámbito de la educación superior. La rápida introducción de la supercomputación en la educación es un ejemplo y representa un desafío que ofrece una gama más amplia de oportunidades para los estudiantes.

La posibilidad de gestionar un gran volumen de datos e información ha llevado a educadores e investigadores a adoptar una visión pedagógica para promover el uso de estas infraestructuras en la enseñanza. En este sentido, la Educación en Computación (CE), es un campo de investigación destacado que enfrenta nuevos desafíos para los que es necesaria la motivación de los estudiantes y es la base de sistemas como el SAIL, un sistema para el aprendizaje adaptativo basado en el interés del alumnado (Aguar et al., 2017a). Actualmente surgen nuevos retos y tareas para el Sistema de Educación Superior (SIE) que requieren de una amplia y completa capacitación, que está condicionada por las limitaciones de la integración de tecnologías en la enseñanza (Suárez, 2007; Wyld, 1996), tal como sucede con las restricciones relacionadas con la supercomputación en los planes de estudio (Holmes y Kureshi, 2015). Actualmente, el uso de supercomputadoras se está convirtiendo en un elemento clave en la mejora de ciertas disciplinas de la educación superior, a pesar de la falta de experiencia y habilidades básicas en este campo (Abuzaghle et al., 2013) y la baja disponibilidad de instalaciones de Supercomputación debido a los altos costes (Fabricius et al., 2005).

La Educación en Supercomputación (SE), como uno de los subcampos de la CE, se considera importante para la innovación, la modernización y el desarrollo tecnológico, contribuyendo a crear una nueva generación de profesionales, empresas y organizaciones relacionadas con la ciencia y la tecnología (Bethel et al., 2011). Esta área también se encarga de una capacitación específica, principalmente para estudiantes de ciencias e ingeniería computacionales y también para estudiantes de diversos campos del conocimiento, que utilizan herramientas computacionales, matemáticas y de ingeniería

para la resolución de problemas (Sung et al., 2003). Actualmente, muchas instituciones educativas de todo el mundo están introduciendo seriamente los conceptos relacionados con la Ciencia de la Computación (CS) en la formación básica de los estudiantes (Aguar et al., 2016), lo que plantea la cuestión de cuán importante es la formación en supercomputación para el futuro profesional de los estudiantes.

La introducción de nuevas tecnologías de la información y la comunicación (TIC) en las sesiones de capacitación de estudiantes de todos los niveles se ha considerado una de las principales herramientas en la modernización de todos los sistemas educativos. Ha contribuido a la evolución de la "sociedad de la información" a la "sociedad del conocimiento" (Martín, 2007), permitiendo formar a los estudiantes, de manera adecuada, para el ejercicio de una gran variedad de profesiones. Esta consideración significa que se podría sugerir la aplicación de los conceptos de CS para el área central de muchas disciplinas de posgrado, donde el uso de una supercomputadora será esencial en la resolución de problemas complejos. Los rápidos cambios en tecnologías informáticas y electrónicas proporcionan desarrollos de alto nivel, especialmente para dar cobertura a las necesidades de la formación en las áreas de la informática biomédica y de salud (Hemanth et al., 2012).

Para las empresas y organizaciones en general, la cualificación tecnológica es imprescindible para asegurar la competitividad corporativa al mejorar la productividad y reducir el consumo de recursos humanos y físicos a través del cambio de entornos de gestión estratégica, basados en el uso de la tecnología de las TIC (Lee y Chang, 2014). En relación con lo anterior, la capacitación en supercomputación representa una actividad de aprendizaje específica que proporciona un enfoque integrado y experimental en el uso de tecnologías para estudiantes de educación superior en contenidos avanzados. Estos contenidos están diseñados para proporcionar experiencias en el análisis de casos reales y para ser utilizados como recursos pedagógicos, al mismo tiempo que aumenta la participación del alumno, haciendo que los estudiantes se sientan involucrados en el

contenido del curso, mejorando el nivel de experiencia técnica y las expectativas con respecto a la tecnología. El nivel de capacitación ofrecido en estos cursos cubre una amplia gama de posibilidades que dependen de la categoría objetivo de los alumnos y cuentan con el apoyo de especialistas en muchos campos de conocimiento que ayudan a los estudiantes a resolver problemas aplicados complejos de computación intensiva. La SE se basa en una amplia variedad de cursos y herramientas, bajo diferentes formas de programas educativos, que permiten el desarrollo de estas habilidades. En general, los cursos vinculados a la Supercomputación difieren de los de capacitación avanzada para especialistas en TI y se pueden clasificar en tres grupos (Hacker, 2010):

- Arquitectura de sistemas: el objetivo es enseñar cómo diseñar, implementar y desplegar una plataforma de supercomputación.
- Aplicaciones en Ciencias de la Computación / Educación en Computación: basadas en la implementación de programas de alto rendimiento en una plataforma de supercomputación.
- Cursos de dominio específico: utiliza la potencia informática de una supercomputadora para resolver problemas específicos de un campo, como la física computacional (Lu et al., 2012), las matemáticas computacionales (Joiner et al., 2006) y la biología computacional (Bader, 2004).

Con respecto a los estudiantes de educación superior, las ventajas de la capacitación en supercomputación son las siguientes:

- El estudiante explora problemas reales a través de la experiencia práctica (Armosky et al., 2007; Bernabé et al., 2014; Farian et al., 2008; Fernández Slezak et al., 2010; Holmes y Kureshi, 2015; Smith y Wolffe, 2007; Stojanovic y Milanovic, 2015) y aprende las funcionalidades más avanzadas de las computadoras, vinculando la teoría con una solución práctica de problemas.
- No siempre es necesario, previamente al curso, tener habilidades en el uso de una Supercomputadora (Ryabinin y Chuprina, 2015).

- La capacitación en supercomputación prepara a los estudiantes en un proceso continuo para obtener un nivel de conocimiento y competencia adecuada para poder utilizar sistemas informáticos complejos (Armosky et al., 2007).

- La capacitación en supercomputación ayuda en el uso de modelos matemáticos (Bergeron et al., 2008; Stainsby et al., 2009; Yasar et al., 2006), lo que permite describir un problema en un campo de investigación específico, para extraer conclusiones y evaluar el desempeño de los problemas propuestos.

- Las simulaciones y el diseño de datos en una Supercomputadora preparan a los estudiantes para el papel de líderes en investigación (Yasar et al., 2006) mediante el uso de sistemas de grandes bases de datos (Abuzaghleh et al., 2013), que cubren, al mismo tiempo, una gran variedad de aplicaciones (Shiflet y Shiflet, 2010).

- La supercomputación contribuye a inventar nuevas herramientas y nuevas filosofías de trabajo, que mejoran el aprendizaje sobre informática (Farian et al., 2008; Kim y Moon, 2019).

La Taxonomía de los objetivos educativos de Bloom (1994) representa una manera conveniente de describir el grado en que los estudiantes entienden y usan conceptos, demuestran habilidades particulares y afectan sus valores, actitudes e intereses. Según esta taxonomía, la capacitación en supercomputación mejora el compromiso del alumno al proporcionar experiencias que les permiten practicar habilidades de aplicación práctica.

Este capítulo presenta, en primer lugar, un análisis de la supercomputación en la educación a través de una revisión de literatura relevante y, en segundo lugar, un estudio de caso que ilustra los beneficios y desafíos de organizar un curso de alta tecnología para estudiantes de educación superior. El objetivo principal es medir la calidad de las sesiones de capacitación mediante el uso de herramientas HPC, analizando la forma en que la integración de la tecnología, en el proceso de enseñanza y aprendizaje, involucra a los estudiantes, proporcionando motivación, actitudes, logros e interacciones (Yang et al.,

2011). Para lograr este objetivo y basado tanto en la revisión de la literatura como en la experiencia de las entidades colaboradoras de este estudio, se desarrolló un cuestionario de encuesta para evaluar la opinión de los estudiantes en relación con la capacitación que han recibido. La retroalimentación obtenida se utilizó como base tanto para validar el diagrama de flujo propuesto como para mejorar las sesiones futuras y, al mismo tiempo, para buscar un mejor rendimiento (Crouch y Mazur, 2001; Denny et al., 2008).

El objetivo principal del capítulo es la descripción de un diagrama de flujo del proceso educativo relacionado con la supercomputación. Los objetivos secundarios son: (1) estudiar los principales campos de aplicación identificados en la revisión de la literatura para estudiantes STEM y no STEM o no informáticos en instituciones de educación superior (HEI) donde se utilizan supercomputadoras, (2) para estudiar los tipos de cursos computacionales, relacionados con la supercomputación, identificados a través de la revisión de la literatura, (3) para proporcionar un estudio relacional que yuxtapone el campo o sector que usa la supercomputación y el tipo de capacitación computacional específica y (4) para analizar cómo de una supercomputadora en un curso práctico, organizado conjuntamente por una universidad y un centro de supercomputación, permite mejorar la capacitación en entornos de educación superior.

Las conclusiones destacan las fortalezas del enfoque experiencial adoptado a través del estudio de caso. Se argumenta que la tecnología, en general, es una forma poderosa de involucrar a los estudiantes con prácticas innovadoras, mejorando así su rendimiento.

El resumen del capítulo es el siguiente: en la Sección 2, discutimos brevemente el material y la metodología seguidos; La Sección 3 presenta, en mayor detalle, una visión general de los resultados obtenidos, con análisis de datos relacionados con el uso de Supercomputadoras en educación. En la sección 4, se describe un estudio de caso de capacitación. Una discusión general de los resultados se describe en la Sección 5. Finalmente, en la Sección 6, se presentan las Conclusiones.

## 5.2.- MATERIALES Y METODOLOGÍA

El alcance del estudio, tanto con respecto a su contenido técnico como educativo, es adecuado para una combinación de métodos: se entrelaza la teoría con los hallazgos de la revisión de la literatura y, al mismo tiempo, se basa en gran medida en el método de estudio de caso.

La base de la población estudiada fue la utilizada en el capítulo 4 (Fernández et al., 2019), que sigue guías relacionadas con las revisiones de literatura (Fink, 2010), con especial atención a las revisiones sistemáticas y está basada en una taxonomía sobre revisiones bibliográficas en educación (Cooper, 1988). Se considera como una verdadera investigación basada en hipótesis, en la que los estudios se seleccionan y combinan mediante el uso de un protocolo predefinido que reduce la subjetividad y la posibilidad de sesgo del investigador (Anderson et al., 2011; Kitchenham, 2004). Los elementos del estudio se filtraron en función de una revisión de revistas, conferencias y libros relevantes, centrándose en el análisis de términos relacionados con la capacitación, la supercomputación y la temática tecnológica en diversos sectores. Las bases de datos consultadas consideradas como los principales repositorios para búsquedas en informática han sido *IEEEExplore Digital Library*, *ACM Digital Library*, *Elsevier ScienceDirect*, *Google Scholar* y *Web of Science*. Como resultado de la primera fase de la búsqueda, se extrajeron 1.911 obras. Una vez que se aplicaron algunos criterios de inclusión / exclusión, 128 elementos se consideraron como los más relevantes. En la segunda fase, se consideraron 8 nuevos trabajos, lo que elevó los elementos a 136. Al final de esta fase 2, se aplicaron criterios de evaluación de calidad y el número final de elementos seleccionados fue 34. El proceso de obtención de los estudios es el que se ha indicado en la Figura 4.1.

El marco metodológico del estudio de caso se basó en el análisis de cursos de Supercomputación llevados a cabo en colaboración entre una Universidad (Universidad de León, España) y un Centro de Supercomputación (SCAYLE). El uso de un diseño de estudio de caso se utiliza para describir y explicar cómo son las experiencias basadas en la

investigación de campo, a fin de establecer su relación con el contexto de la vida real (Yin, 2009). En general, los resultados del estudio de caso ofrecen una gran cantidad de ideas y amplían las experiencias del lector (Merriam, 1998) en su construcción de conocimiento (Stake, 1994).

Los cursos incluidos en el estudio de caso han demostrado que mejoran tanto la competencia informática como los procesos de enseñanza y aprendizaje, introduciendo a los estudiantes al uso de varias herramientas tecnológicas y prácticas de integración de tecnología, para usar efectivamente supercomputadoras.

Con el fin de ofrecer una mejor calidad en las acciones formativas, la implementación de ambos cursos siguió los Estándares Nacionales de Tecnología Educativa para Docentes (NETS) de la Sociedad Internacional de Tecnología en Educación (ISTE). Los estándares ISTE (2008) se clasificaron de la siguiente manera:

- (a) facilitar e inspirar el aprendizaje y la creatividad de los estudiantes,
- (b) diseñar y desarrollar experiencias y evaluaciones de aprendizaje de la era digital,
- (c) modelar el trabajo y el aprendizaje de la era digital,
- (d) promover y modelar ciudadanía y responsabilidad digital, y
- (e) comprometer el crecimiento profesional y el liderazgo.

El principal interés de este capítulo radica en analizar cómo las experiencias proporcionadas por la capacitación en supercomputación mejoran el rendimiento de los estudiantes de educación superior y no en el análisis de aspectos técnicos más profundos en relación con el HPC.

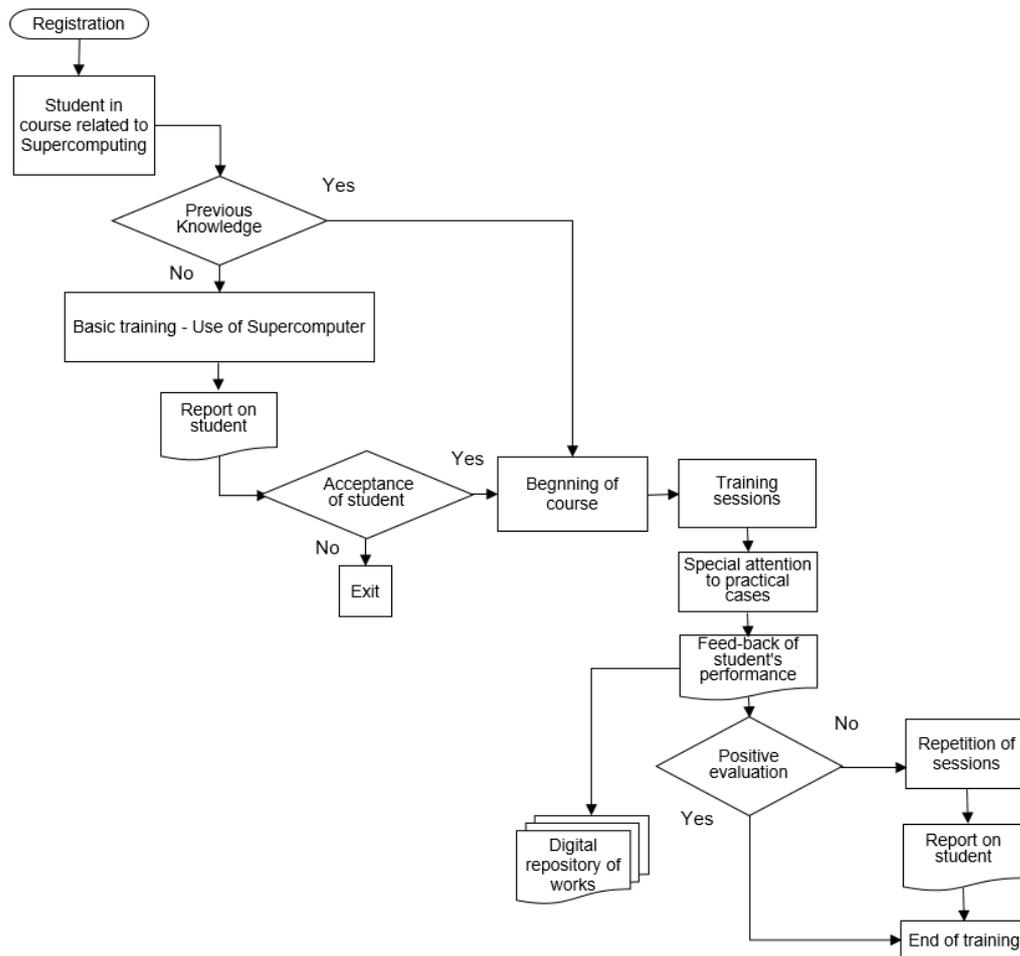
### **5.3.- RESULTADOS**

Como se mencionó en apartados anteriores, el objetivo principal de este capítulo es presentar la descripción de un diagrama de flujo del proceso educativo relacionado con la Supercomputación, cuya definición se completó una vez que se desarrollaron algunos

objetivos parciales de este primer estudio: (1) el estudio de los principales campos de aplicación de la Supercomputación para los estudiantes, (2) la identificación del tipo de cursos computacionales relacionados con la Supercomputación en la revisión de la literatura y (3) el estudio relacional del campo o sector que usa la Supercomputación y el tipo de capacitación computacional específica. En la siguiente sección, se describe el cuarto objetivo secundario: el análisis basado en el estudio de caso de un curso práctico.

La Figura 5.1 describe un diagrama de flujo en el que se analiza el desarrollo de un curso de Supercomputación, desde el análisis previo del perfil de los estudiantes hasta su cierre. El proceso se basa en la descripción de las tareas realizadas por las entidades que han colaborado en el estudio de caso presentado, estrechamente relacionado con los resultados de la revisión de la literatura. El primer paso del proceso, una vez que el estudiante ha sido registrado, es verificar su nivel de conocimiento previo, especialmente en el uso de sistemas operativos tipo Unix como Linux. Se ofrecerá una formación específica de estos sistemas, pero, en el caso de que el estudiante no posea el nivel mínimo de conocimiento necesario, no se le permitirá iniciar el curso, al considerar que no obtendrá un adecuado aprovechamiento. Una vez que comience el curso, se prestará especial atención a los casos prácticos, ya que es una de las principales experiencias que buscan los estudiantes: saber cómo aplicar la supercomputación en la solución de sus tareas. En este proceso, se proporcionan comentarios sobre el rendimiento de los estudiantes en aras de mejorar el desarrollo del curso, así como un repositorio digital en el que se almacenan los casos a ser utilizados como ejemplos en futuras sesiones de capacitación. Antes del final del curso, se evalúa el aprovechamiento de los estudiantes y, si es necesario, se repiten aspectos en los que pueda haber dudas. Cuando los profesores consideran que los alumnos han entendido todos los conceptos, tanto teóricos como prácticos, el curso finaliza, ofreciendo la posibilidad de repetir algunas sesiones.

**Figura 5.1: Diagrama de flujo de un proceso de capacitación relacionado con la supercomputación**



Fuente: elaboración propia

Las siguientes secciones describirán la solución propuesta para cada objetivo parcial, descritos anteriormente: Campos de aplicación, tipo de cursos de contenido computacional y estudio relacional que compara el campo de conocimiento que utiliza la supercomputación y el tipo de capacitación computacional específica.

### 5.3.1.- Campos de aplicación

Para una mejor comprensión de los detalles sobre los estudios de cada campo, analizado en la revisión, se utilizó la nomenclatura internacional de la Unesco para agrupar los campos de conocimiento más relevantes en el uso de supercomputadoras, de acuerdo con el contenido de cada artículo.

La Tabla 5.1 muestra los campos que usan supercomputadoras en virtud de los datos encontrados en la revisión de la literatura. La columna "Categoría" muestra un total de 52 referencias relacionadas con campos de conocimiento. En la columna "Especialidad", se incluye una referencia más específica del tipo de actividad dentro de cada categoría.

**Tabla 5.1: Campos de conocimiento que utilizan la supercomputación**

Categoría	Especialidad	Número de artículos
Matemáticas (4)	Matemáticas y análisis numérico	1
	Problemas de control	1
	Control automatizado	1
	Diseño automatizado	1
Astronomía y Astrofísica (3)	Astrofísica	1
	Exploración del espacio	1
	Ingeniería aeroespacial	1
Física (8)	Fenómenos físicos	1
	Procesos físicos	3
	Dinámica termal y de fluidos	4
Química (6)	Química Computacional, Ingeniería Química y Química Estructural	6
Ciencias de la vida (8)	Biología	3
	Bio-informática	2
	Genómica and proteómica	2
	Simulación molecular	1
Ciencias de la tierra y del espacio (9)	Meteorología y predicción meteorológica	5
	Ciencia de los materiales	3
	Predicción para reforestación de bosques	1
Ciencias de la Tecnología (12)	Procesamiento de imagen	2
	Diseño de fármacos	2
	Fabricación de vehículos (terrestres / anfibios / aéreos): diseño estructural, diseño de motores, modelado financiero y test de choques	1

	Ingeniería	2
	Microelectrónica	2
	Tecnología de sensores	1
	Nanotecnología	1
	Medios de comunicación	1
Ciencias Económicas (1)	Economía y Finanzas	1
Lingüística (1)	Lingüística	1

Fuente: elaboración propia

### 5.3.2.- Tipo de cursos de contenido computacional relacionados con la supercomputación

En el ámbito de la misma revisión de literatura, se realizó un análisis de los cursos con contenido computacional relacionados con la supercomputación.

Como se ve en la Tabla 5.2, la columna "Categoría" muestra un total de 97 referencias detectadas a través de la revisión de la literatura, relacionadas con diferentes tipologías de cursos computacionales identificados. En la columna "Especialidad", se describe una referencia más específica al contenido del curso.

**Tabla 5.2: Cursos computacionales relacionados con la supercomputación**

Categoría	Especialidad	Número de artículos
Programación (16)	Programación paralela	10
	Programación multidispositivo	1
	Técnicas de programación para supercomputadores	1
	Programación paralela y distribuida	2
	Programación multinúcleo y evaluación de desempeño	1
	Optimización de código	1
	Clúster Computing	3

Computación (17)	Cloud Computing	2
	Multicore, Multicore computing y Multicore clustering	1
	Computación paralela	3
	Computación paralela de alta capacidad	2
	Computación paralela y distribuida	2
	Tecnología HPC (hardware)	1
	Computación escalable	1
	Cálculo matricial	1
	Computación heterogénea	1
Software (6)	Aplicaciones	1
	Desarrollo de usuarios	1
	Herramientas de desarrollo de ingeniería de software	1
	Ingeniería de software	1
	Diseño de software para sistemas complejos y paralelos	1
	Software de paralelismo.	1
Sistemas y arquitecturas (17)	Arquitectura de sistemas paralelos	8
	Sistemas distribuidos y paralelo	4
	Administración de sistemas	1
	Sistemas paralelos	1
	Sistemas High Performance Computing	1
	Sistemas de computación científica	1
	Sistemas operativos	1
Gestión de datos (14)	Herramientas de visualización de datos científicos y matemáticos	4
	Base de datos distribuida y paralela	1
	Gestión de workflows	1
	Análisis de datos / Post-proceso	1
	Partición de datos	1
	Extracción de datos	1
	Bases de datos distribuidas y paralelas	2
	Big Data	2
Otros (28)	Algoritmos paralelos	6
	Optimización	1

	Clúster y Grid Middleware	1
	Tecnologías de red	1
	Modelado y simulación	6
	Cuestiones operacionales	1
	Análisis de rendimiento	1
	Verificación y Validación	1
	Dominios	1
	Paquetes de aplicaciones HPC	1
	Recursos científicos	1
	Bibliotecas Paralelas	1
	Aplicaciones científicas de paralelismo	1
	Balanceo de carga	1
	Checkpoints	1
	Introducción a los aceleradores	1
	Infraestructura de red	1
	Procesamiento paralelo	1

Fuente: elaboración propia

### **5.3.3.- Estudio relacional que compara el campo de conocimiento que utiliza la supercomputación y el tipo de capacitación computacional específica**

El estudio relacional se basó en la identificación de 12 artículos que mencionan, al mismo tiempo, un campo de conocimiento y un tipo de curso de computación, lo que permite realizar comparativas. El número inicial de artículos seleccionados se redujo, porque no todos mencionaron datos de ambas variables al mismo tiempo. Para evaluar la relación entre la variable "campo de conocimiento" y la variable "tipo de formación", se han seguido los criterios de la UNESCO para definir dos dimensiones: Ciencias (incluidas todas las asignaturas relacionadas con la ciencia y las relacionadas con la economía y la lingüística) e Ingeniería (incluye artículos relacionados con la subdisciplina de Matemáticas, Astronomía y Astrofísica, Química y Física). Para lograr los objetivos propuestos, se ha realizado la construcción de dos modelos de regresión múltiple, siendo los campos

dependientes los campos de Ciencias e Ingeniería mencionados anteriormente. En todos los casos, las variables independientes fueron los diferentes tipos de cursos computacionales desarrollados, descritos en la Tabla 5.2.

Es necesario enfatizar que, en las relaciones buscadas para el desarrollo de los modelos, el número de referencias o frecuencia en que se mencionan las variables en la selección final de artículos detallados en el índice, se analiza como se refleja en las Tablas 5.1 y 5.2. El número final de artículos a utilizar en el estudio fue de 12, en los que se encontró coincidencias de referencias, tanto en el campo del conocimiento, como en el tipo de curso mencionado.

Los resultados obtenidos en relación con los parámetros de cada modelo, con un nivel de significación del 95%, se muestran en la Tabla 5.3:

**Tabla 5.3: Modelos creados según el criterio “campo de conocimiento”**

Modelo	Criterio	Predictores	Coefficientes ( $\beta$ )	Error típico	Estadístico t
Modelo 1	Ciencias	Intercepto	0	N/A	N/A
		Programación	2,34	0,94	2,49
		Computación	-0,10	0,75	-0,14
		Software	-0,24	2,56	-0,09
		Sistemas y Arquitecturas	-0,68	0,77	-0,89
		Gestión de datos	-1,31	1,34	0,98
		Otros	0,16	0,51	0,31
Modelo 2	Ingeniería	Intercepto	0	N/A	N/A
		Programación	0,01	1,09	0,02
		Computación	0,32	0,87	0,36
		Software	-0,33	2,97	-0,11
		Sistemas y Arquitecturas	0,04	0,89	0,05
		Gestión de datos	0,55	1,54	0,35
		Otros	0,79	0,59	1,32

Fuente: elaboración propia

En el caso de la capacitación en el campo de las Ciencias (Modelo 1), el parámetro de capacitación en Programación es el que se relaciona positivamente en mayor proporción para explicar el aumento en la capacitación de cursos computacionales. En el Modelo 2, el parámetro que mejor explica el aumento en el número de cursos en Ingeniería es el relacionado con Otros (refiriéndose a formación sobre algoritmos, equilibrio de carga, simulaciones, etc.).

Para evaluar la validez y el grado de correlación que existe en los modelos, así como su significación estadística, en la tabla 5.4 se puede observar el análisis de varianzas (ANOVA). Los coeficientes de correlación entre las variables dependientes e independientes de cada modelo, en la Tabla 5.5, muestra el porcentaje para el cual las variables se explican entre sí.

**Tabla 5.4: Resultados detallados de ANOVA para cada modelo**

Modelo 1

	Grados de libertad	Suma de cuadrados	Promedio de cuadrados	F	Valor crítico de F
Regresión	6	60,09	10,02	6,75	0,0268
Residuo	6	8,91	1,48		
Total	12	69			

Modelo 2

	Grados de libertad	Suma de cuadrados	Promedio de cuadrados	F	Valor crítico de F
Regresión	6	21,08	3,51	1,76	0,274
Residuo	6	11,92	1,99		
Total	12	33			

Fuente: elaboración propia

A partir de los datos presentados anteriormente, se puede observar que solo el Modelo 1 puede considerarse como globalmente válido al satisfacer la condición de que el valor crítico F sea menor que 0.05 y, por lo tanto, la hipótesis nula sería rechazada.

Adicionalmente al análisis ANOVA, podemos ver, en la Tabla 5.5 el coeficiente  $R^2$  ajustado que permite analizar cuál de los modelos propuestos explica mejor la correlación entre las variables. En el Modelo 1, donde la variable dependiente es el campo "Ciencias", tiene un porcentaje de correlación más alto, lo que explica el grado de variación de la variable dependiente en función de las variables independientes seleccionadas. El valor correspondiente del Modelo 2 está por debajo de 0.5 en la escala inferior de los parámetros que miden este coeficiente y, por lo tanto, no explica adecuadamente la correlación como ocurre en el Modelo 1.

**Tabla 5.5: Detalle de la correlación de los modelos**

Regresión Estadística		
	Modelo 1	Modelo 2
	Ciencias	Ingeniería
Coefficiente de correlación múltiple	0,93	0,80
Coefficiente de determinación $R^2$	0,87	0,64
$R^2$ ajustado	0,60	0,17
Error típico	1,21	1,41
Número de observaciones	12	12

Fuente: elaboración propia

#### 5.4.- ESTUDIO DE CASO

Esta sección describe el estudio de caso basado en los cursos completados por estudiantes relacionados con la supercomputación. Los cursos analizados fueron organizados por la Universidad de León y el Centro de Supercomputación de Castilla y León (SCAYLE), y se llevaron a cabo durante 2016. Los objetivos, en ambos casos, eran proporcionar conceptos fundamentales y avanzados mediante el uso de sistemas de bases de datos de grandes volúmenes de datos mediante el uso de una Supercomputadora. Los cursos se pueden clasificar en la categoría "Cursos de dominio específico" y pertenecían al campo de la biotecnología, al ser la especialidad más demandada de este tipo de cursos en las instituciones analizadas.

En la descripción de los estudios de caso, la realización de un análisis previo es de gran importancia. De hecho, prácticamente todos los hallazgos de la revisión de la literatura previa son aplicables al desarrollo de las acciones formativas descritas.

Los nombres de los cursos fueron: "Curso práctico de introducción al uso de la supercomputación aplicada al análisis de datos de RNA-Seq - 2a edición" (C1) y "Curso práctico sobre el uso de la supercomputación aplicada a la metagenómica y la genómica comparativa "(C2). Las características más relevantes de los cursos se resumen en la Tabla 5.6.

**Tabla 5.6: Caracterización de los cursos (datos significativos)**

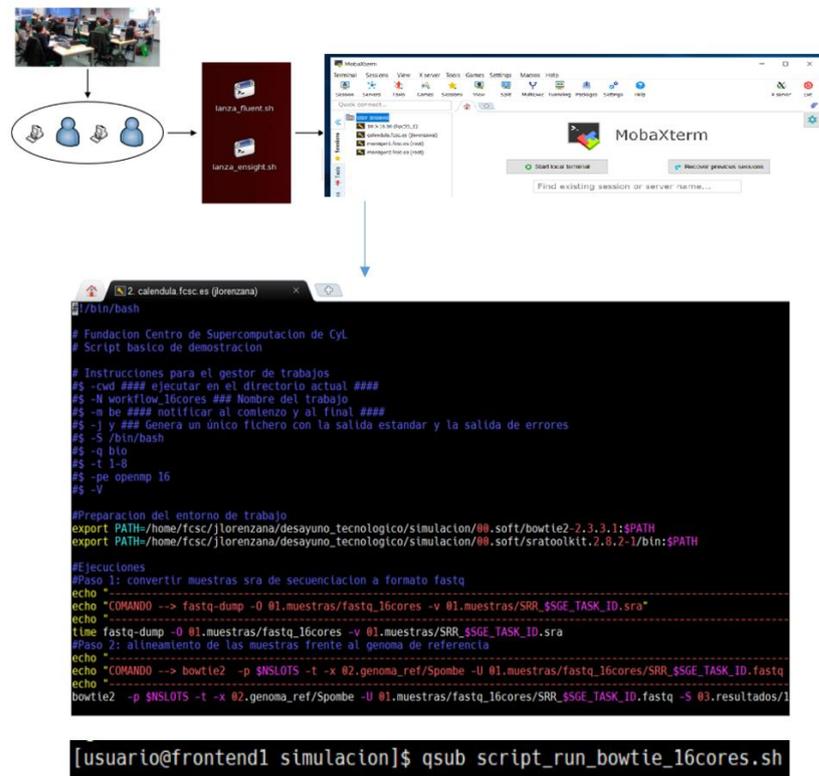
	C1	C2
Organizador	Universidad de León	Centro de Supercomputación de Castilla y León
Fechas	Julio de 2016	Octubre de 2016
Número de estudiantes	16	14
Duración de curso	28 horas	
Realización de seminario previo de Linux para un mejor uso de los recursos de supercomputación.	Sí	
Número de estudiantes que han asistido al seminario previo	16 (100 %)	12 (86 %)
Tipo de participantes en los cursos	Investigadores y/o estudiantes de educación superior interesados en estudios genómicos, profesionales de la informática, biólogos y/o biotecnólogos relacionados con el diagnóstico genético y estudiantes universitarios de posgrado (con calificaciones técnicas en el campo experimental) y, en general, cualquier persona relacionada con el tema de investigación, innovación y desarrollo.	
Objetivos	Gestión e interpretación de datos de la expresión génica global de la secuenciación de próxima generación utilizando una supercomputadora.	

Fuente: elaboración propia

Los programas de formación relacionados con la Supercomputación normalmente los llevan a cabo especialistas altamente cualificados con amplia experiencia en sistemas y aplicaciones HPC y, por lo general, también tienen habilidades para el uso de las prestaciones de las supercomputadoras para una amplia gama de problemas informáticos de diferentes campos del conocimiento. Estos especialistas enseñan a los estudiantes la forma de optimizar el software utilizado para trabajar de manera eficiente y, a través de nuevas estrategias de enseñanza, basadas en la mejora de procesos pedagógicos y aplicando un enfoque educativo correcto, les facilita la adopción de la tecnología basada en la supercomputación para su formación, por medio de la personalización de los planes de aprendizaje. En todo caso, los docentes también tienen que adaptarse, según el conocimiento previo y los objetivos de aprendizaje individuales descritos en la literatura, a estudiantes con una gran variedad de antecedentes y debilidades formativas previas específicas (Aguar et al., 2017b). Para una mejor organización y calidad, los cursos se diseñaron, desarrollaron y probaron utilizando la infraestructura de un clúster de una supercomputadora real y un software debidamente programado y optimizado para generar resultados rápidos para resolver problemas complejos.

El análisis de los gráficos que se muestran en las Figuras 5.2 y 5.3 proporciona información sobre cómo ha sido el uso de la infraestructura de supercomputación durante el desarrollo de los cursos. En la Figura 5.2, podemos ver el diagrama de cómo se utilizó la supercomputadora en las sesiones C1 y C2. Una vez que los estudiantes completan la fase teórica, comienzan a evaluar casos prácticos, relacionados con la especialización del curso, enviando trabajos a la supercomputadora. Tienen una retroalimentación inmediata de la simulación que ejecutan y pueden ver directamente la calidad y la eficiencia del uso de estas infraestructuras en la optimización del rendimiento de un proceso, mejorando el aprendizaje.

**Figura 5.2: Descripción general del trabajo de los estudiantes en los cursos de supercomputación**

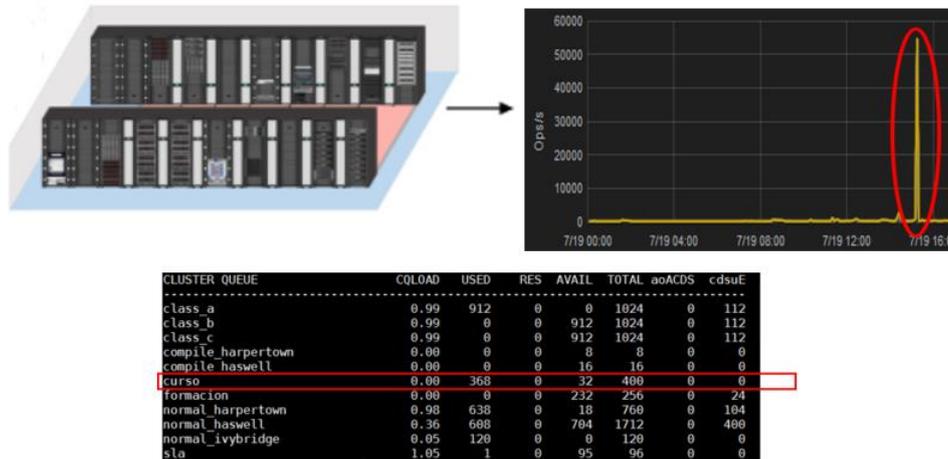


Fuente: elaboración propia

Como una parte básica y esencial de los cursos, los estudiantes realizan simulaciones computacionales utilizando datos reales. Una herramienta gráfica se encarga de controlar la carga de la infraestructura para proporcionar la mejor eficiencia a los usuarios en el uso de la supercomputadora (Figura 5.3). El gráfico muestra un pico marcado con un círculo rojo, que representa las operaciones por segundo en el sistema de almacenamiento. El detalle resaltado corresponde al momento exacto en que se procesaron estos datos en las sesiones prácticas y muestra el pico de uso cuando los estudiantes acceden a los datos almacenados en el disco para comenzar los cálculos. El gráfico presentado se basa en los sistemas de monitoreo de SCAYLE (Grafana, ver web <https://grafana.com>). La imagen de la pantalla en la parte inferior muestra una lista completa de los trabajos que se estaban ejecutando en la Supercomputadora en el mismo momento, destacando los servidores asignados al curso en el administrador de colas de clúster. Esto se obtiene del comando

qstat del administrador de trabajos SGE, que muestra todas las colas creadas y su estado (núcleos disponibles, núcleos totales, núcleos utilizados, carga promedio, etc.).

**Figura 5.3: Detalle de herramientas gráficas durante el desarrollo de un curso**



Fuente: elaboración propia

#### 5.4.1.- Participantes y recopilación de datos

Ambos cursos fueron diseñados para estudiantes con antecedentes diversos y generalmente están formados por un máximo de 12 a 20 participantes, a fin de permitir una educación más personalizada. El número y las características de los estudiantes se muestran en la Tabla 5.7 a continuación.

**Tabla 5.7: Perfil de los alumnos**

Perfil del estudiante	C1	C2
Doctor con experiencia	2	2
Post-doctorado	2	2
Investigador	11	10
Estudiante de Máster	1	

Fuente: elaboración propia

Una de las claves de una investigación es tener claro su propósito y alcance. En este sentido, la implementación de métodos para la recolección y análisis de datos, como procesos

iterativos, son esenciales para todo tipo de evaluaciones y por lo tanto aplicable al caso de este estudio, con el fin de completar el análisis cuantitativo que permita obtener información para analizar los resultados. Para cumplir con este fin, se proporcionó a los estudiantes, al final de las sesiones de ambos cursos, un cuestionario que constaba de 10 ítems que combinaba preguntas de tipo cerrado y abierto.

La confiabilidad del cuestionario se probó previamente en SCAYLE, donde esta herramienta se utiliza como un estándar habitual para evaluar la calidad de la formación impartida en sus propios programas docentes. El cuestionario proporciona, mediante el uso de preguntas clave específicas de evaluación, la evidencia necesaria para emitir juicios sobre la calidad de los cursos. La consistencia de los ítems del cuestionario fue evaluada por expertos que también contribuyeron al diseño y adaptación de los instrumentos de recolección de datos.

#### 5.4.2.- Resultados de la encuesta

El cuestionario de evaluación incluye una primera parte que evalúa los aspectos introductorios del curso, una segunda parte que requiere una evaluación de los maestros y, en la última parte, una consideración final de todo el curso. La Tabla 5.8 incluye el texto completo de las preguntas que se incluyeron en la encuesta, y muestra los resultados de los estudiantes que respondieron a la misma. Los ítems pueden clasificarse en tres grupos: (i) aspectos introductorios (Q1 a Q5); (ii) evaluación de los docentes (Q6 y Q7); (iii) satisfacción general y consideraciones finales (P8 a P10).

**Tabla 5.8: Resultados de la encuesta**

Pregunta		C1 N=15	C2 N=13
P1	¿El seminario introductorio de Linux fue de utilidad?	Sí: 15	Sí: 13
P2	Valoración media del perfil del profesor del seminario de introducción, perfil biotecnólogo (1-5)	4,62	4,92

P3	Valoración media del perfil del profesor de seminario de introducción, ingeniero informático (1-5)	4,60	4,40
P4	¿Considera adecuado el material de enseñanza?	Sí: 12	Sí: 11
		No contestadas: 3	No contestadas: 2
P5	¿Se ha difundido adecuadamente el curso?	Sí: 14	Sí: 13
		No: 1	No: 0
P6	¿Cambiarías algo sobre el seminario?	No: 5	No: 8
		No contestadas: 5	No contestadas: 4
	- Reducir la parte técnica del seminario de inicio.	1	1
	- Incluir más aspectos prácticos en las sesiones.	3	
	- Tener en cuenta que los estudiantes tienen conocimientos básicos de computación.	1	
	- Rebajar un poco el nivel del curso	1	
	- Ampliar el contenido del curso	3	
P7	Valoración media de los profesores investigadores (1-5)	3,59	4,46
P8	¿Ha cubierto el curso sus expectativas?	Sí: 12	Sí: 10
		No contestadas: 3	No contestadas: 3
P9	¿Ha sido atractiva la metodología del curso?	Sí: 12	Sí: 10
		No contestadas: 3	No contestadas: 3
P10	Evaluación del curso (1-10)	8,83	9,69

Fuente: elaboración propia

El análisis de las respuestas de cada grupo mencionado anteriormente es el siguiente:

—Aspectos introductorios. El seminario introductorio fue considerado muy útil por todos los estudiantes en ambos cursos, por lo que es esencial solventar la falta de conocimiento sobre los aspectos necesarios para usar una supercomputadora. En este sentido, la calificación tanto de los profesores de perfil de biotecnólogos, como de ingenieros informáticos en los seminarios introductorios fue muy alta, con calificaciones cercanas al máximo. Finalmente, la mayoría de los estudiantes consideran que los otros aspectos introductorios, como el material proporcionado en el curso y la difusión, fueron adecuados.

—Evaluación de los docentes: para determinar la percepción del alumno sobre los cursos, se planteó la pregunta Q6, relacionada con posibles cambios en el curso. El número de preguntas sin respuesta fue alto, pero, en cualquier caso, se propuso un contenido más práctico y una mayor extensión y profundidad. En el caso de la evaluación de los docentes especializados (P7), podemos ver que los docentes del curso C2 fueron evaluados mejor, probablemente porque C1 fue un curso muy intensivo, con el mismo número de sesiones y el doble de docentes, haciendo la tarea de enseñanza más difícil.

—Gestión general y consideraciones finales. Las preguntas P8 a P10 proporcionan la evaluación final del curso. La mayoría de los estudiantes considera que el curso cumplió con sus expectativas (P8) y que el método de enseñanza fue atractivo (P9). La evaluación final de ambos cursos fue muy alta: 8,83 para C1 y 9,69 para C2.

## **5.5.- DISCUSIÓN**

Este estudio primero analiza los resultados obtenidos en la revisión de la literatura y, luego, los del estudio de caso. Al final de esta sección, se proporcionará información sobre la relación de ambos estudios.

### **5.5.1.- Supercomputación en el proceso educativo**

El aumento de las infraestructuras de supercomputación ha sido paralelo a la complejidad de la investigación y los problemas que enfrenta la sociedad actual y se ha desarrollado para obtener mejores resultados en diversos campos del conocimiento y sectores de actividad. Actualmente, la simulación numérica para obtener resultados de investigación se practica ampliamente, tanto en el ámbito académico como en la industria en general, para resolver problemas en un corto período de tiempo. Por esta razón, la capacitación en estas tecnologías es esencial para un desempeño adecuado de los profesionales que trabajan en diversos tipos de campos y actividades.

Para la consolidación de la supercomputación en el proceso educativo, es esencial invertir en la capacitación de personal competente en el uso adecuado de las supercomputadoras (Coveney et al., 2012). La capacitación en supercomputación mejora los resultados de los estudiantes STEM y no STEM, o no informáticos, en una institución de educación superior (HEI), entre otros. Con la ayuda de muchas herramientas, como un software nuevo y fácil de usar, la simulación por computadora se hace accesible a un grupo más amplio de personas.

Para implementar la supercomputación, es esencial la necesidad de cooperación entre profesionales de múltiples habilidades y de diferentes antecedentes, lo que supone un valor agregado importante, que debe basarse en dos aspectos interrelacionados: (1) diferentes perfiles de una variedad de campos de conocimiento en diferentes competencias en STEM y no STEM; (2) diferentes experiencias en la integración de tecnología en el sistema de Educación Superior.

Nuestros resultados sugieren, a través de los resultados vistos en la revisión de la literatura, que incluir supercomputadoras en los procesos de aprendizaje de una gran variedad de campos ha brindado muchas oportunidades para los estudiantes.

### **5.5.1.1.- Resultados**

Muchos factores, como las limitaciones en el plan de estudios, como los rápidos cambios en el campo de la supercomputación en general, o la sofisticación de los modelos matemáticos, influyen en el uso de las instalaciones de supercomputación.

Este estudio proporciona tres resultados principales relacionados con la capacitación de Supercomputación de vanguardia, que responden a los objetivos planteados al comienzo de este estudio y que se resumen a continuación.

- Objetivo 1: mediante el análisis de los 52 artículos sobre el tratamiento de diferentes campos relacionados con el uso de supercomputadoras, se detectó un número importante de campos que se consideran no STEM o no computacionales, como una gran cantidad de especialidades, que son la base de muchos procesos industriales.
- Objetivo 2: se encontró que muchos de los cursos de computación que se describen en la literatura, influyen en las actividades relacionadas con la supercomputación. Estos cursos mejoran claramente el rendimiento de los usuarios de estas infraestructuras, al mejorar los procesos en los que se basa la investigación, lo que permite una mejor interpretación de los resultados.
- Objetivo 3: Según el estudio relacional analizado, el Modelo 1, vinculado a los ítems vinculados con el campo de la Ciencia, muestra que la programación es la variable estrecha relacionada con este campo de conocimiento.

### **5.5.1.2.- Limitaciones**

Durante el curso de la revisión de la literatura, la mayor dificultad encontrada fue la heterogeneidad en el tipo de artículos, y también en el contenido relacionado con la tipología de la capacitación. Otra limitación del estudio fue el hecho de que, debido a la naturaleza de la información, no se centró en analizar la capacitación relacionada con la Supercomputación como un elemento del proceso de enseñanza-aprendizaje, por lo que

sería necesario complementar el estudio con otros relacionados con el uso de computadoras de alto rendimiento como nuevas herramientas didácticas.

Finalmente, otra limitación está relacionada con la metodología utilizada en los modelos de regresión, que se basa en un enfoque correlacional. Los resultados logrados nos permiten hablar de una cierta relación entre el campo del conocimiento y el tipo de curso computacional especializado, pero que no necesariamente significa una causalidad entre estas variables. Para un mayor desarrollo de esta investigación, serán necesarios estudios basados en diseños experimentales para confirmar el efecto de variables independientes en el desarrollo de la capacitación computacional en los campos del conocimiento.

También esperamos que el presente estudio aliente a los investigadores e instituciones en general a contribuir al desarrollo de acuerdos que aborden los desafíos de usar una supercomputadora como parte fundamental de la investigación de muchos campos, mediante el uso de una infraestructura de supercomputación real y la inclusión en el currículo académico de estudiantes STEM y no STEM de materias relacionadas con este tema.

## **5.5.2.- Estudio de caso**

### **5.5.2.1.- Resultados**

Este estudio de casos brinda la oportunidad de examinar cómo los procedimientos de capacitación en un entorno tecnológico proporcionan y facilitan la participación de los estudiantes en la mejora de su desempeño (Coates, 2005). Las cuestiones planteadas deberían ser relevantes para proporcionar nuevas herramientas a los estudiantes de educación superior, complementando así sus habilidades en el uso de la tecnología. Destacamos que, a pesar de algunos casos descritos en los que los estudiantes participan en el diseño curricular (Carey, 2013), actualmente, los contenidos en HPC son prácticamente testimoniales en el currículo académico de la educación superior (Holmes y Kureshi, 2015; Munshi et al., 2006; Smith y Wolffe, 2007).

La capacitación en supercomputación apoya el rendimiento de los estudiantes y mejorar las actividades de aprendizaje al proporcionar una plataforma que ayuda en la búsqueda de casos reales en muchos campos. En el futuro, el diseño de cursos tecnológicamente avanzados debe realizarse con el uso de una supercomputadora, avanzando así con respecto a la capacitación ofrecida y siguiendo nuevos métodos pedagógicos.

Sin embargo, aunque estos cursos han recibido, en general, comentarios favorables y evaluaciones positivas, algunos estudiantes expresaron la necesidad de mejorar algunos aspectos en el futuro, como la falta de más tiempo para sesiones prácticas y explicaciones más profundas. Además, el alto nivel del curso en términos de habilidades computacionales específicas que se presume requiere atención, porque algunos estudiantes declararon tener un fondo computacional de bajo nivel utilizando una Supercomputadora de manera óptima. Las preguntas más importantes que pueden considerarse para futuros cursos son las siguientes:

—Los estudiantes consideran esencial el apoyo del seminario introductorio. En algunas preguntas abiertas, incluso expresan la creencia de que necesitarían más capacitación en algunas áreas, antes del comienzo del contenido principal del curso para optimizar su rendimiento.

—Incrementar el apoyo de los técnicos de los Centros de Supercomputación, para mejorar y ampliar la capacidad de usar las infraestructuras para probar los aspectos prácticos de los cursos, algo muy demandado por los estudiantes.

—Mejorar los aspectos didácticos planificados de las clases, para garantizar el mejor rendimiento posible de los profesores.

—La implementación del diseño de una interfaz fácil de usar debe llevarse a cabo, para facilitar la interacción humano-computadora con una experiencia intuitiva durante el uso de una supercomputadora.

### **5.5.2.2.- Limitaciones**

Este estudio tiene ciertas limitaciones que deben considerarse al interpretar sus resultados y conclusiones. En primer lugar, un mayor número de participantes sería más adecuado porque, debido al número relativamente pequeño de participantes de ambos cursos (30, de los cuales 28 respondieron a la encuesta), los resultados deben interpretarse con precaución. En segundo lugar, los cursos considerados en el estudio fueron impartidos por diferentes instructores, en áreas específicas, y los perfiles de los estudiantes también fueron diferentes, lo que complica el estudio de una población homogénea. Para minimizar los efectos de estas limitaciones, serían esenciales niveles similares de educación y experiencia de los docentes, y un estrecho seguimiento de la implementación del contenido en ambos cursos.

En el desarrollo del trabajo, ha sido posible verificar cómo la mayoría de los aspectos analizados en la revisión de la literatura se han observado en los cursos analizados en el estudio de caso. Así, aspectos como la falta de conocimiento previo en el uso de una supercomputadora, la notable mejora en la calificación de los estudiantes, o la disponibilidad efectiva de infraestructuras de supercomputación para llevar a cabo casos prácticos de simulación de la realidad, han permitido realizar acciones para mejorar la formación ofrecida.

## **6.- ANÁLISIS DEL DESARROLLO Y LA VALIDEZ PREDICTIVA DEL CUESTIONARIO DE CAPACITACIÓN RELACIONADO CON LA SUPERCOMPUTACIÓN**

---

### **RESUMEN**

En este capítulo se presenta el proceso de desarrollo y validación de un cuestionario para la elaboración de una encuesta de autoevaluación que permita examinar el contenido tecnológico y pedagógico, así como los problemas de conocimiento relacionados con el uso de técnicas de supercomputación. Se ha observado que existe una clara necesidad de proporcionar herramientas para los formadores con el objetivo de aumentar su conocimiento de las técnicas de enseñanza y mejorar las actividades de aprendizaje de los estudiantes.

En este estudio, se identificaron dos factores, a través de una encuesta basada en una muestra de 97 asistentes a cursos basados en el uso de una supercomputadora en varios campos, organizados en el Centro de Supercomputación de Castilla y León - SCAYLE (España). Los datos fueron recolectados usando un cuestionario de 18 ítems, que fue previamente validado por un comité de expertos y probado en una población piloto para perfeccionar la versión final. Este cuestionario se deriva del modelo de Conocimiento pedagógico tecnológico y de contenido (TPACK) y tiene un valor alfa de Cronbach aceptable para la confiabilidad de la escala. Además, el análisis factorial exploratorio ha indicado un buen ajuste para una estructura de dos factores. Sobre la base de estos resultados, se propuso un modelo de cuestionario para la evaluación de los enfoques de enseñanza y aprendizaje en el campo de la supercomputación. El cuestionario podría usarse para mejorar las sesiones de formación en relación a futuras investigaciones sobre la aplicación de TPACK en contextos de supercomputación y también para buscar implicaciones que sugieren el desarrollo de otras encuestas de TPACK específicas.

Palabras clave: supercomputación; Formación; Cuestionario

## **6.1.- INTRODUCCIÓN**

En este capítulo se presentan recursos para mejorar el aprendizaje en entornos de supercomputación, basados en un estudio exploratorio, a través de una adaptación del modelo de conocimiento pedagógico, tecnológico y de conocimiento (TPACK) en la percepción de estudiantes asistentes a cursos relacionados con la supercomputación. Este tipo de estudiante necesita entornos flexibles para apoyar el uso de la tecnología informática en la formación.

Las posibilidades de los estudiantes, de acuerdo con el enfoque constructivista, que prevalece en las sociedades actuales, se mejoran mediante el uso de nuevas tecnologías que les permiten crear y mejorar su propio conocimiento, adaptando el proceso de aprendizaje a sus necesidades (Área, 2005) y, por lo tanto, mejorar la motivación y los resultados de este proceso de aprendizaje. Durante los últimos años, ha habido un creciente interés en cómo la tecnología de la información y la comunicación (TIC) se ha utilizado para mejorar el aprendizaje, influir en la naturaleza de cada disciplina, para proporcionar nuevas formas y oportunidades de apoyar el aprendizaje y un mayor acceso a la educación (OECD, 2015). En la literatura se han descrito muchas metodologías en el campo educativo que analizan la integración de las TIC en la educación, con el fin de mejorar el proceso de aprendizaje de los estudiantes. De la conjunción entre la teoría educativa constructivista y la aparición de nuevas TICs surge el modelo: Conocimiento pedagógico, tecnológico y de contenido o TPACK, que es un modelo teórico basado en la mejora de la calidad de la enseñanza mediante el uso eficiente de las TIC para la construcción de conocimiento por parte de los alumnos. Para poder alcanzar el éxito es esencial que los formadores tengan las habilidades necesarias en tecnología, disciplina y pedagogía (De Benito y Lizana, 2012). La investigación desarrollada en este capítulo se basa en el estudio de Koehler y Mishra (2005) y las conclusiones de Schmidt et al (2009), que abordan este fenómeno mediante el desarrollo de la primera encuesta TPACK.

Para la creación del cuestionario definitivo, se hizo una propuesta basada en el análisis de teorías y conocimientos previos de 52 manuscritos relacionados con la capacitación y la

tecnología, vinculados con las dimensiones del modelo TPACK descrito por Koehler y Mishra en 2005. El escenario elegido para evaluar el modelo fueron cursos en los que grupos de estudiantes usaban una supercomputadora como base de la formación.

En el estudio se ha considerado un enfoque experimental para validar las premisas del estudio. El diseño del cuestionario incorporó características que ayudaron a extraer conclusiones para un mejor desempeño.

El resumen del capítulo es el siguiente: en la Sección 6.2, discutimos brevemente la base teórica de TPACK, a través de una revisión de la literatura existente relevante que constituye nuestro contexto de trabajo; La Sección 6.3 describe la metodología para el diseño e implementación del cuestionario, basada en la brecha de investigación y el propósito de nuestro estudio; La Sección 6.4 presenta, en mayor detalle, una visión general de los resultados obtenidos, con análisis de datos de la encuesta de una base de datos de asistentes a cursos basados en HPC. Una discusión general de los resultados y sus implicaciones, y también las sugerencias para la investigación de seguimiento, se discuten en las conclusiones e investigaciones futuras que se describen en la Sección 6.5. En la sección 5.6 se presenta la Discusión.

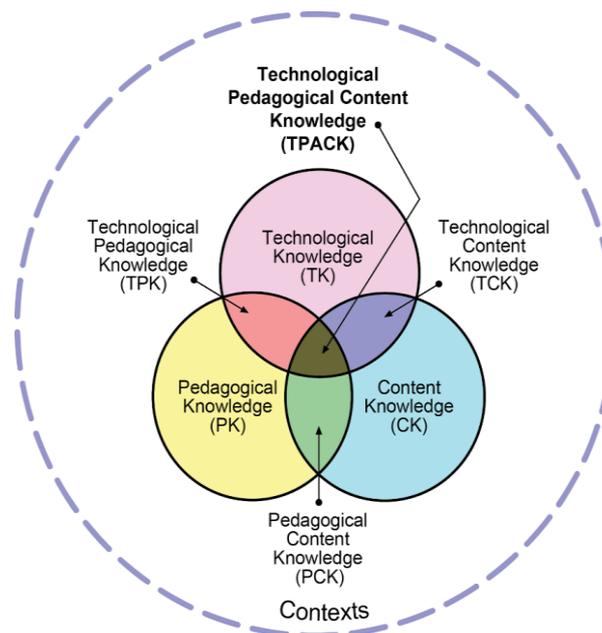
## **6.2.- REVISIÓN DE LITERATURA**

El cuestionario propuesto, en el que se centra la investigación, se basa en el modelo TPACK mencionado anteriormente, cuyo marco de desarrollo se ha basado en la construcción de Shulman (1986) del Conocimiento del Contenido Pedagógico (PCK) para incluir posteriormente el conocimiento de la tecnología, así como el conocimiento pedagógico. Los trabajos de Grossman (1990), De Vicente (1994), Pierson (2001) y Angeli y Valadines (2005), describen cómo se utiliza el conocimiento de los formadores para mejorar la pedagogía para el aprendizaje y la eficiencia de la enseñanza.

El modelo TPACK es un marco que intenta comprender y describir la naturaleza del conocimiento requerido por los formadores para la integración efectiva de la tecnología de las prácticas pedagógicas en un entorno complejo de aprendizaje tecnológico en un enfoque

constructivista y representa una de las metodologías más importantes para la integración de la tecnología en educación. TPACK consta de tres componentes: conocimiento de contenido, conocimiento pedagógico y conocimiento tecnológico. La integración de estos componentes y elementos producirá áreas comunes en las que se puede encontrar dos o más elementos que interactúan entre sí (Koehler y Mishra, 2005). TPACK debería ser idealmente una síntesis del conocimiento tecnológico (TK), el conocimiento pedagógico (PK) y el conocimiento del contenido (CK) de los docentes. Más importante aún, se argumenta que el modelo TPACK necesita reconocer las interacciones entre el conocimiento pedagógico tecnológico (TPK), el conocimiento del contenido tecnológico (TCK) y el conocimiento del contenido pedagógico (PCK). El modelo TPACK se representa en la Figura 6.1:

**Figura 6.1: Modelo TPACK**



Fuente: <http://tpack.org> (Reproducido con permiso del publicador, © 2012 by tpack.org).

La cantidad y calidad de las investigaciones y los artículos publicados relacionados con el modelo TPACK han aumentado constantemente en la última década (Chai et al., 2013a; Wu, 2007). La investigación empírica se ha centrado tradicionalmente en el servicio (Graham, et al., 2009a) y los servicios previos a educadores (Young et al., 2012; Abbitt,

2011; Baran et al., 2011; Albion, et al., 2010; Schmidt et al., 2009; Graham, et al., 2009b) para saber cómo perciben los formadores el TPACK en su práctica docente. Durante este período, varios autores han descrito que el modelo TPACK, basado en la encuesta de Schmidt et al. (2009), era un instrumento confiable y que se podría adaptar al estudio de la percepción de los estudiantes en diferentes condiciones: estudiantes de primaria (Gutiérrez et al., 2015), estudiantes universitarios de segundo y tercer año (Chai et al., 2013b), estudiantes de institutos de ingeniería (Fawad y Manarvi, 2014), estudiantes de educación física (Semiz e Ince, 2012), estudiantes universitarios (Medina et al., 2014; Muñoz-Repiso y Del Pozo, 2016), estudiantes de doctorado (Smith y Bhattacharya, 2015) y estudiantes-docentes en programas educativos (Wang, 2013).

### **6.3.- METODOLOGÍA**

La metodología de este estudio se basó en las diferentes actuaciones desarrolladas en cada una de las cinco fases definidas: 1) pool de ítems, para obtener una versión inicial de posibles preguntas, 2) validación por expertos, para proporcionar opiniones de alto nivel sobre el cuestionario, 3) una prueba piloto, para examinar y refinar los ítems del cuestionario propuesto por expertos, 4) una encuesta final: adaptación de la encuesta modelo TPACK, y 5) un estudio final.

#### **6.3.1.- Pool de ítems.**

Para el desarrollo del pool de ítems iniciales utilizamos el análisis de teorías y conocimientos previos encontrados en manuscritos relacionados con la formación y la tecnología, vinculados con las dimensiones del modelo TPACK descrito por Koehler y Mishra en 2005.

La revisión de la literatura se realizó mediante la consulta de bases de datos científicas (Scopus, Web of Science, Science Direct, etc.), en las que el contenido consultado se basó en cuestionarios en el área de las TIC, lo que permitió la recopilación de información adecuada para el mejor desarrollo de la escala del cuestionario basado en el modelo

TPACK. Esta revisión se utilizó para encontrar elementos que permitieran desarrollar una versión inicial del cuestionario para ser enviado al Comité de Expertos, con el propósito de llevar a cabo un estudio piloto con el que se obtendrían datos para el desarrollo de un cuestionario definitivo.

Para garantizar la representatividad y relevancia de los ítems generados, el modelo TPACK y los detalles de cada uno de sus componentes fueron revisados continuamente realizando las siguientes preguntas: ¿A qué contribuye? ¿El contenido tiene que ver con el modelo TPACK? ¿Es relevante esta contribución? Una vez que se realizó la revisión de la literatura sobre formación especializada en tecnologías de la información, se construyeron los ítems y se agruparon de acuerdo con su significado en cada uno de los cinco tipos de conocimiento descritos en el modelo TPACK. Para la escritura formal de las preguntas, se tomaron en consideración las pautas que reflejan los textos psicométricos, que se refieren a la estructura, el lenguaje y la longitud de los ítems, basados en dos principios: el procedimiento general de construcción del cuestionario (escalas de tipo, por ejemplo) y un conjunto de consejos más o menos bien intencionados sobre cómo escribir buenos artículos (por ejemplo, Edwards, 1957; Dawes, 1975; Morales, 1988 o Foddy, 1996).

De los 52 artículos más relevantes de la revisión, se realizó una selección de 53 ítems posibles, de modo que el primer borrador del cuestionario, enviado al Comité de Expertos, se adaptó mejor al modelo TPACK.

Con respecto al formato de respuesta propuesto al Comité de Expertos para su validación, se adoptó un modelo de respuesta tipo Likert, con cinco opciones (calificadas de uno a cinco): 1 = totalmente en desacuerdo, 2 = en desacuerdo, 3 = indeciso, 4 = de acuerdo, 5 = totalmente de acuerdo. Este tipo de escalas son fáciles de usar e interpretar para el investigador y muy intuitivas para los estudiantes. Además, el estudio de análisis factorial que se pretende desarrollar para analizar la confiabilidad del cuestionario es más efectivo cuando la escala Likert consta de al menos 5 categorías (Ferrando y Anguiano-Carrasco, 2010; Frías-Navarro y Pascual, 2012).

### 6.3.2.- Validación por expertos.

Para el proceso de validación del contenido del cuestionario, se solicitó la colaboración de un Comité de Expertos, con el objetivo de obtener opiniones bien fundadas que permitan identificar las fortalezas y debilidades del instrumento, lo que ha facilitado un análisis profundo que permitió modificar, integrar y eliminar partes del cuestionario inicialmente propuesto. El juicio experto es una estrategia con ventajas considerables y la validación del contenido de un cuestionario a través del juicio experto se ha definido como la opinión informada de personas con experiencia en el tema, reconocida por otros como expertos calificados en ella y que pueden dar información, evidencia, juicios y evaluaciones para el desarrollo de un estudio (Escobar y Cuervo, 2008, p.29). Otros autores, (Cabero et al., 2013, p.14), subrayan que la información detallada sobre el tema a estudiar, derivada de la opinión de los expertos, permite obtener calidad en las respuestas, un alto nivel de profundidad en la evaluación de tema y facilidad de poner en acción, entre otros, la determinación del conocimiento de contenidos y temas, que pueden ser difíciles, complejos y novedosos o poco estudiados. El trabajo de los expertos se realizó de manera completamente independiente y en su desarrollo propusieron cambios en la redacción de los ítems, mejorando así la precisión de las expresiones, evaluaron la adecuación de los ítems con respecto al estudio y propusieron mejoras en la puntuación. La Tabla 6.1 muestra las características de los participantes expertos.

**Tabla 6.1: Detalle del Comité de Expertos**

<b>Entidad</b>	<b>Nivel académico</b>	<b>Puesto</b>
CSIC	Doctor	Investigador
Fundación FISABIO	Doctor	Investigador
Universidad de León	Pre-Doctor	Investigador
Centro de Supercomputación de Castilla y León (SCAYLE)	Doctor	Técnico
Centro de Supercomputación de Castilla y León (SCAYLE)	Ingeniero	Técnico

Instituto Karolinska Estocolmo	Doctor	Investigador
Universidad de León	Doctor	Investigador/Profesor
Universidad de León	Doctor	Investigador
COMPUTAEX	Ingeniero	Técnico
COMPUTAEX	Doctor	Técnico

Fuente: elaboración propia

Una vez que los expertos validaron el contenido mediante métodos cualitativos, el siguiente paso en el desarrollo de la escala se basó en análisis cuantitativos, utilizados para determinar la calidad de los ítems para vincular conceptos abstractos con indicadores observables y medibles (Wynd et al., 2003). El objetivo de estas medidas cuantitativas era evaluar la adecuación de los ítems revisados por el Comité de Expertos. Para ello, se analizaron los siguientes índices:

- Índice de validez de contenido (IVC): es un promedio de la validez de contenido de todos los elementos seleccionados en el paso anterior. Con respecto a la interpretación de este índice, cuando se emplean entre 9 y 40 expertos, los valores mínimos adecuados son .78 y .29, respectivamente y deberían alcanzar un grado máximo de acuerdo (.99) cuando se utilizan 7 o menos expertos. Lynn (1986) proporciona una interpretación similar, quien establece el valor mínimo del índice, teniendo en cuenta el número de expertos involucrados y el número de expertos que consideran el ítem como relevante. Análogamente, otros investigadores han propuesto puntos de corte que miden, al mismo tiempo, el número de elementos evaluados, la consistencia interna de las escalas de evaluación e incluso las implicaciones prácticas de los instrumentos de medición (Crocker, et al., 1988). Se pueden encontrar ejemplos de aplicación directa de este índice en numerosos trabajos aplicados a diferentes áreas, tal es el caso de autores como Bazarganipour, et al. (2012) en el campo clínico, Castle (2008) en el lugar de trabajo y Yeun y ShinPark (2006) al analizar la validez transcultural de un instrumento.

- Coeficiente de validez de contenido (CVC): este coeficiente permite evaluar el grado de acuerdo de los expertos (se recomienda la participación de entre tres y cinco expertos) con respecto a cada uno de los diferentes ítems y el instrumento en general (Hernández-Nieto, 2002). Para esta determinación, luego de la aplicación de una escala Likert de cinco alternativas, se calcula el promedio obtenido en cada uno de los ítems y, en base a esto, se calcula el CVC para cada elemento. Con respecto a su interpretación, Hernández-Nieto (2002) recomienda mantener sólo aquellos ítems con un CVC superior a .80, aunque algunos criterios menos estrictos establecen valores superiores a .70 (Balbinotti, 2004). El trabajo de Balbinotti, et al. (2007), presenta la aplicación de este método al traducir y adaptar una escala centrada en un contexto financiero.

### **6.3.3.- Prueba piloto**

La siguiente fase consistió en una prueba piloto para examinar y refinar los ítems del cuestionario propuesto por los expertos a fin de mejorar tanto la validez como la confiabilidad del instrumento desarrollado. Un estudio piloto se define como "una prueba a pequeña escala de los métodos y procedimientos que se utilizarán a gran escala" (Porta, 2008) y también puede denominarse "prueba piloto" o "prueba previa". Es un paso inicial esencial en un proyecto de investigación, aplicable a todo tipo de estudios y representativo de una buena investigación (Hazzi y Maldaon, 2015). Una prueba piloto implica una prueba a pequeña escala de una población particular y se considera un diseño cuasiexperimental. En este documento, la base de la prueba piloto fue un curso de 5 días relacionado con el uso de la supercomputación en genética. Participaron 18 estudiantes que asistieron al curso, completando la versión borrador del cuestionario basado en la evaluación previa de expertos. Los datos fueron recolectados a través de un cuestionario de encuesta en línea.

La prueba a realizar mediante un borrador del cuestionario antes de la recolección de datos de la encuesta final es muy importante. Este estudio piloto nos permitió mejorar la versión final del cuestionario. Las pruebas preliminares y las pruebas piloto pueden ayudar a identificar algunas preguntas que no tienen sentido para los participantes u otros problemas

con el cuestionario que podrían dar lugar a respuestas sesgadas. Los aspectos que se evaluarán mediante una prueba piloto para mejorar la calidad del cuestionario son, al menos, los siguientes:

- Las reacciones de los encuestados pueden verse desde diferentes ángulos.
- Análisis del tiempo apropiado para responder el cuestionario.
- Aceptación de las preguntas formuladas (descubrir si la repetitividad o la redundancia es molesta o descubrir temas que no cubren las expectativas de lo que el participante esperaba).
- Nivel de cooperación de los participantes en la encuesta.
- Descubrimiento de posibles errores en el cuestionario.
- Verificación del procedimiento general de la muestra: cómo los participantes siguen las instrucciones, el tiempo total, etc.
- Evaluación del logro de los objetivos de la formación.
- Evaluación de diferentes instrumentos y software para el procesamiento y análisis de datos para una mejor interpretación de los resultados.

Una vez que los participantes completaron el borrador del cuestionario, fue necesario verificar la confiabilidad. Para este propósito, el alfa de Cronbach es la medida más común (Field, 2009), porque es una medida de la consistencia interna de una escala, es decir, describe el grado en que todos los elementos de una escala miden lo mismo. Una revisión de la literatura muestra que la de Cronbach es la herramienta más utilizada en muchas disciplinas, especialmente cuando los estudios abordan escalas de respuestas múltiples como son las de tipo Likert, utilizadas principalmente en la investigación psicológica empresarial, industrial y social.

#### **6.3.4.- Adaptación de la encuesta modelo TPACK**

Para obtener el cuestionario final y un número razonable de variables apropiadas para el tamaño estimado de la población en estudio, se procedió a agrupar las variables de las dimensiones del cuestionario resultantes de la validación de la prueba piloto no relacionada con la tecnología (PK, CK y PCK), dejando el resto de las variables relacionadas con la tecnología que se consideraron válidas para continuar con la siguiente fase del estudio. Para obtener las variables finales que se aplicarían en la encuesta a ser enviada a la población final, se aplicó el modelo Stepwise. Este método, sobre la base de un criterio específico, consiste en un procedimiento de pasos para ajustar modelos de regresión en el que la elección de variables predictivas se lleva a cabo mediante un procedimiento automático y se considera para sumar o restar del conjunto de variables explicativas.

Varios investigadores (Huberty, 1989; Huberty, 1994; Thompson, 1989) destacaron el uso de análisis escalonados como algo común. Según Thompson (1989, p146), "los métodos analíticos por etapas pueden estar entre las prácticas de investigación más populares empleadas tanto en la investigación sustantiva como en la de validez".

#### **6.3.5.- Estudio final**

La recopilación formal de datos comenzó en abril y finalizó en mayo de 2018, con el cuestionario de encuesta final refinado de la prueba piloto, que abarcaba una población total de 370 estudiantes de la Base de Datos del Centro de Supercomputación de Castilla y León (SCAYLE), que habían estado participando en cursos y sesiones de formación del Centro. En total, se obtuvieron 97 respuestas válidas a la encuesta con una tasa de respuesta del 26,21%. El análisis de los datos de la encuesta fue generalmente cuantitativo en su enfoque, utilizando el programa SPSS 24.0, apropiado para el trabajo con estadísticas descriptivas respecto a los perfiles personales y profesionales del encuestado. Se utilizaron estadísticas descriptivas para presentar medidas de tendencia central, como la media y la desviación estándar.

Para la validación del constructo, se realizó un análisis factorial de los componentes principales, en el que se consideraron aquellos factores con valores propios superiores a uno. En términos de rotación de factores, se utilizaron tanto la rotación de Varimax como la de Oblimin, ya que varios autores lo han recomendado (Gable y Wolf, 1986; Kline, 1999) como una buena forma de comparar. Dado que los resultados de ambas soluciones fueron similares, presentamos e interpretamos los resultados de la rotación Varimax, que es el tipo más común ya que es ortogonal (Nunnally y Bernstein, 1994: 536), la opción preferida (Rennie, 1997), aunque no necesariamente el mejor, de hecho, algunos autores como Costello y Osborne (2005) recomiendan la rotación oblicua (Oblimin directo).

Los supuestos para la aplicación del análisis factorial se verificaron utilizando el índice Kaiser-Meyer-Olkin (KMO) y la prueba de esfericidad de Bartlett. Se considera adecuado un valor entre 0.5 y 1.0 para el índice KMO y significativo para  $p < 0.05$  para la prueba de esfericidad de Bartlett. La confiabilidad de la encuesta se exploró determinando el coeficiente alfa de Cronbach, interpretando valores  $> 0.7$  como buena confiabilidad.

## **6.4.- ANÁLISIS DE DATOS**

El análisis de datos se desarrolló en cada uno de los apartados definidos en la metodología que permitieron de forma gradual el desarrollo de la investigación. Los detalles de cada una de las fases son los siguientes:

### **6.4.1.- Pool de ítems**

El detalle de la búsqueda llevada a cabo para la definición de los ítems del borrador del cuestionario para ser enviado a la Comisión de Expertos fue el siguiente:

- Manuscritos comprendido entre los años 2007 a 2017.
- Los términos de búsqueda fueron: calidad del cuestionario, “curso ict”, cuestionario, “curso TIC”.
- Primeros resultados: 515 obras (artículos, manuscritos, etc.).

- Selección final: 52 manuscritos.

- Artículos identificados: 53 artículos.

Para la selección de los elementos identificados para el futuro cuestionario, se realizó un estudio en profundidad de los manuscritos seleccionados comparando su contenido con la escala original TPACK (Schmidt et al, 2009). Los ítems extraídos durante el proceso de selección se compararon con los de la escala original, extrayendo elementos con una redacción similar a la misma para ser propuestos para el estudio de expertos. En el APÉNDICE II, se describe la selección de 53 elementos (sin considerar los de información personal de los participantes en la encuesta como edad, género, etc.) con base en la revisión de la literatura.

#### **6.4.2.- Comité de Expertos**

La versión resultante del cuestionario inicial que se remitió a los expertos consistió en un cuestionario de 53 ítems, cada uno con una escala Likert de 5 puntos (ver APÉNDICE II) con información sobre género, edad, campo de conocimiento y conocimiento previo, con el objetivo de, una vez analizado por los expertos, se desarrollara una versión revisada y para ser aplicada en una población piloto.

La evaluación de expertos consistió, en primer lugar, en una opinión sobre la redacción de cada ítem, puntuando del 1 al 5 su grado de idoneidad y proponiendo cambios para una mejor comprensión por parte de la población final. Además de esto, los resultados del Índice de Validez del Contenido (CVI) y el Coeficiente de Validez del Contenido (CVC) determinan la eliminación o aceptación de los ítems previamente propuestos. El resultado final, una vez que se han calculado tanto el CVI como el CVC, muestra que se eliminaron 15 ítems, lo que resultó en un constructo final de 38 ítems. Los expertos no modificaron la primera sección del cuestionario (género, edad, campo, nivel académico, institución y acceso a tecnología y capacitación), como se puede observar en los Apéndices III y IV.

### 6.4.3.- Prueba piloto

Los participantes de la prueba piloto fueron 18 estudiantes que asistieron a un curso relacionado con el uso de la supercomputación en el campo de la genómica, organizado por el Centro de Supercomputación de Castilla y León (SCAYLE). En relación con el número de participantes en la muestra de la prueba piloto, Billingham et al. (2013) consideran que, aunque todos los estudios deben tener una justificación del tamaño de la muestra, otros estudios no necesitan tener un cálculo del tamaño de la muestra. En cualquier caso, la conclusión de Billingham et al. (2013) es que un cálculo formal del tamaño de la muestra para estudios piloto puede no ser apropiado. Baker (1994) considera del 10 al 20% del tamaño de la muestra principal como un porcentaje adecuado para realizar un estudio piloto; por lo tanto, el número de participantes en este estudio se considera adecuado. Por lo expuesto, se considera válida la muestra utilizada en la prueba realizada.

En los resultados de la prueba piloto (Tabla 6.2), se observa un valor promedio alto y un valor de varianza bajo (medida de dispersión que representa la variabilidad con respecto a la media), lo que indica, en un primer análisis, que los datos pueden ser aceptables para el estudio.

**Tabla 6.2: Sumario de estadísticas de los ítems**

	Media	Mínimo	Máximo	Rango	Máximo / Mínimo	Varianza	Número de elementos
Media de los elementos	4,092	2,556	4,833	2,278	1,891	,305	38
Varianza de los elementos	,762	,147	1,987	1,840	13,511	,234	38
Covarianzas entre elementos	,254	-,376	1,582	1,958	-4,209	,042	38
Correlaciones entre elementos	,371	-,508	,925	1,433	-1,822	,059	38

Fuente: elaboración propia

En términos de confiabilidad de la escala, la Tabla 6.3 muestra el valor del Alfa de Cronbach, que es un modelo de consistencia interna basado en el promedio de correlaciones entre los ítems. El valor es alto; de hecho, algunos autores muestran que valores superiores a .70 son adecuados (Gliem y Gliem, 2003).

**Tabla 6.3: Estadísticas de fiabilidad**

Alfa de Cronbach	Alfa de Cronbach basada en elementos estandarizados	Número de elementos
,950	,957	38

Fuente: elaboración propia

En la Tabla 6.4, las estadísticas de los ítems muestran que, una vez analizados los resultados, se procedió a eliminar los ítems cuyo índice de homogeneidad (correlación de elementos totales corregidos) fue menor a 0.3 (Stewart et al, 1992), lo que contribuye a un aumento del Alfa de Cronbach. Para refrendar el proceso, se revisó el criterio de algunos autores que consideran que se puede aceptar un valor a partir de 0.3 en el índice de homogeneidad (Kline, 1999).

**Tabla 6.4: Estadística totales de los Ítems**

	Escala promedio si el elemento se elimina	Escala de la varianza si el elemento se elimina	Corrección total de elementos correlacionados	Correlación múltiple al cuadrado	Alfa de Cronbach si el elemento se elimina
TK1	152,9444	365,232	,452	.	,950
TK2	152,7778	364,889	,392	.	,951
TK3	152,3889	35166,0	,534	.	,950
TK4	152,5556	364,614	,545	.	,949
TK5	152,4444	353,791	,585	.	,949
PK1	151,3889	368,369	,542	.	,949
PK2	151,8889	359,516	,698	.	,948
PK3	151,7778	374,771	,224	.	,952
PK4	151,5556	361,908	,714	.	,948
PK5	151,0556	382,526	,180	.	,951
PK6	150,8889	375,752	,534	.	,949
PK7	151,5556	363,438	,667	.	,948
PK8	151,3889	366,369	,670	.	,948

CK1	151,2778	365,507	,727	.	,948
CK2	150,6667	385,059	,078	.	,951
CK3	151,1667	371,324	,552	.	,949
CK4	151,3333	364,235	,802	.	,948
TPK1	151,2778	360,801	,744	.	,948
TPK2	151,5000	357,088	,686	.	,948
TPK3	151,0000	370,000	,814	.	,948
TPK4	151,2778	359,859	,670	.	,948
TPK5	150,7222	379,977	,373	.	,950
TPK6	151,0556	366,526	,717	.	,948
TPK7	150,8889	374,575	,486	.	,949
TCK1	151,0556	372,761	,482	.	,949
TCK2	151,0000	370,000	,584	.	,949
TCK3	151,2222	374,889	,285	.	,951
PCK1	150,8889	369,046	,628	.	,949
PCK2	151,3333	358,824	,670	.	,948
PCK3	151,5556	362,026	,658	.	,948
PCK4	151,5000	352,618	,666	.	,948
PCK5	151,3889	361,899	,691	.	,948
TPACK1	151,2222	368,418	,682	.	,948
TPACK2	150,8889	369,046	,727	.	,948
TPACK3	151,0556	371,467	,531	.	,949
TPACK4	151,1667	364,735	,808	.	,948
TPACK5	151,2222	364,301	,847	.	,947
TPACK6	151,2222	366,536	,887	.	,948

Fuente: elaboración propia

En relación con el análisis planteado siguiendo las premisas propuestas en líneas anteriores, se decidió eliminar los ítems PK3, PK5, CK2 y TCK3 por tener bajos índices de homogeneidad, realizando nuevamente el análisis de confiabilidad.

Como se observa en la Tabla 6.5, los resultados del nuevo análisis arrojan un valor promedio alto y un valor de varianza bajo, similar a los resultados en la Tabla 6.3, lo que indica nuevamente que los datos pueden ser aceptables para el estudio, una vez que las variables mencionadas anteriormente fueron eliminadas.

**Tabla 6.5: Sumario de estadísticas de los ítems**

	Media	Mín.	Máy.	Rango	Máximo / Mínimo	Varianza	Número de elementos
Media de los elementos	4,065	2,556	4,778	2,222	1,870	,316	34
Varianza de los elementos	,771	,183	1,987	1,804	10,857	,230	34
Covarianzas entre elementos	,296	-,131	1,582	1,712	-12,100	,036	34
Correlaciones entre elementos	,431	-,111	,925	1,037	-8,303	,042	34

Fuente: elaboración propia

Una vez que los elementos mencionados en el primer análisis han sido eliminados, el Alfa de Cronbach es más alto que el cuestionario inicial (Tabla 6.6), por lo que podemos inferir que el cuestionario tiene buena consistencia interna y más alta que antes de eliminar los ítems. Un constructo final de 34 ítems se considera un instrumento válido que, no obstante, debe adaptarse para su uso con la población final a ser considerada, no observándose ningún elemento adicional que pueda ser eliminado. En la tabla 6.7, el análisis de las estadísticas totales de los ítems muestra que las variables resultantes tienen todas ellas un índice de homogeneidad mayor de 0.3, cumpliendo el criterio de ser admisibles (Kline, 1999).

**Tabla 6.6: Estadísticas de fiabilidad**

Alfa de Cronbach	Alfa de Cronbach basada en elementos estandarizados	Número de elementos
,955	,963	34

Fuente: elaboración propia

**Tabla 6.7: Estadísticas totales de los ítems**

	Escala promedio si el elemento se elimina	Escala de la varianza si el elemento e elimina	Corrección total de elementos correlacionados	Correlación múltiple al cuadrado	Alfa de Cronbach si el elemento se elimina
TK1	135,6667	337,765	,449	.	,955
TK2	135,5000	336,500	,409	.	,956
TK3	135,1111	328,458	,541	.	,955
TK4	135,2778	336,095	,573	.	,954
TK5	135,1667	326,618	,586	.	,955
PK1	134,1111	341,399	,519	.	,954
PK2	134,6111	333,075	,673	.	,953
PK4	134,2778	335,036	,697	.	,953
PK6	133,6111	347,546	,549	.	,954
PK7	134,2778	336,448	,652	.	,953
PK8	134,1111	338,222	,691	.	,953
CK1	134,0000	338,000	,726	.	,953
CK3	133,8889	343,399	,558	.	,954
CK4	134,0556	337,114	,788	.	,953
TPK1	134,0000	332,824	,764	.	,952
TPK2	134,2222	330,654	,665	.	,953
TPK3	133,7222	342,448	,806	.	,953
TPK4	134,0000	332,235	,679	.	,953
TPK5	133,4444	351,673	,387	.	,955
TPK6	133,7778	339,124	,711	.	,953
TPK7	133,6111	347,310	,459	.	,954
TCK1	133,7778	344,536	,498	.	,954
TCK2	133,7222	341,742	,605	.	,954
PCK1	133,6111	341,193	,636	.	,953
PCK2	134,0556	332,056	,656	.	,953
PCK3	134,2778	335,507	,631	.	,953
PCK4	134,2222	325,124	,676	.	,953
PCK5	134,1111	334,340	,696	.	,953
TPACK1	133,9444	340,291	,702	.	,953
TPACK2	133,6111	341,075	,740	.	,953
TPACK3	133,7778	343,242	,549	.	,954
TPACK4	133,8889	337,399	,801	.	,953
TPACK5	133,9444	336,879	,844	.	,952
TPACK6	133,9444	338,997	,886	.	,952

Fuente: elaboración propia

#### **6.4.4.- Encuesta final: adaptación de la encuesta modelo TPACK**

Para la elaboración del cuestionario final se debe considerar una serie de ítems adaptados a la población objetivo. En el caso del presente estudio, debemos considerar que, debido a sus especificaciones y características especiales, un número relativamente reducido de estudiantes asiste a cursos de supercomputación. De hecho, la encuesta se entregó a los estudiantes de la Base de Datos del Centro de Supercomputación de Castilla y León (SCAYLE), compuesta por 370 estudiantes que asistieron a cursos durante la última década, y finalmente se obtuvieron un total de 97 respuestas válidas a la encuesta, lo que representa una tasa de respuesta del 26,21%.

Para determinar el número mínimo de personas necesarias por elemento de la encuesta y, por lo tanto, asegurar que el estudio llevado a cabo sea riguroso, se han analizado varios estudios similares en los que se considera que una proporción de 5 personas por elemento como válida. En este sentido, varios autores (Carretero y Pérez, 2005; MacCallum et al., 1999; Osborne y Costello, 2004; Bryant y Yarnold, 1995; Nunnally, 1978; Hair, et al., 1995; Plichta et al., 2013) consideran muestras de entre 5 y 10 individuos por ítem como un número válido para evaluar la validez de contenido de una escala y contribuir de manera adecuada a la estabilidad de la solución factorial.

Una vez que se obtuvieron los resultados de la prueba piloto, el siguiente paso fue utilizar un método de enfoque gradual denominado “Stepwise”, para reducir el número de variables en la versión final del cuestionario, de modo que se pueda obtener una media de 5 participantes por ítem. Para el análisis gradual descrito, los ítems obtenidos como resultado de la prueba piloto se agruparon en dos dimensiones: primero, los elementos relacionados con la pedagogía, el conocimiento y el contenido y, segundo, los relacionados con la tecnología.

El método paso a paso “Stepwise” es muy popular en la investigación educativa (Whitaker, 1997) y constituye una forma de seleccionar las mejores variables que se utilizarán en el modelo mediante el filtrado de variables. Muchos autores han utilizado este método de

reducción variable por pasos (Wagner et al 2007; George, 2000 y Murtaugh, 2009) en el desarrollo de sus trabajos. Mediante la aplicación del método “Stepwise”, se eliminaron algunos elementos relacionados con la tecnología. Los resultados se analizan en las Tablas 6.8, 6.9 y 6.10. En la tabla 6.8 se detalla el estudio de los diferentes modelos según se va considerando la inclusión de variables, de forma que se puede observar que no se elimina ninguna variable.

**Tabla 6.8: Variables de entrada y eliminadas <sup>a</sup>**

Modelo	Variables de entrada	Variables eliminadas	Método
1	TPACK6	.	Por pasos (Criterios: Probabilidad de F para ingresar $\leq$ , 050, Probabilidad de F para eliminar $\geq$ , 100).
2	TPK1	.	Por pasos (Criterios: Probabilidad de F para ingresar $\leq$ , 050, Probabilidad de F para eliminar $\geq$ , 100).
3	TCK2	.	Por pasos (Criterios: Probabilidad de F para ingresar $\leq$ , 050, Probabilidad de F para eliminar $\geq$ , 100).
4	TPK6	.	Por pasos (Criterios: Probabilidad de F para ingresar $\leq$ , 050, Probabilidad de F para eliminar $\geq$ , 100).

a. Variables dependientes: sumaPK\_CK\_PCK

Fuente: elaboración propia

El resumen de la bondad del modelo que se puede observar en la tabla 6.9 muestra que el modelo 4 es el que presenta mejores valores, con lo que se utilizará de base para el cuestionario final. En este análisis estadístico, el Modelo 4 con TPACK6, TPK1, TCK2, TPK6 como variables independientes, representa el modelo más representativo, basado en el valor del parámetro  $R^2$ .

**Tabla 6.9: Modelos resumidos<sup>e</sup>**

Modelo	R	R <sup>2</sup>	R <sup>2</sup> Ajustado	Error estándar de la estimación	Estadísticas de cambio					Durbin - Watson
					R cuadrado	F	df1	df2	Sig. F	
1	,762 <sup>a</sup>	,580	,554	6,06442	,580	22,115	1	16	,000	
2	,826 <sup>b</sup>	,683	,640	5,44587	,102	4,841	1	15	,044	
3	,878 <sup>c</sup>	,770	,721	4,79731	,088	5,330	1	14	,037	
4	,916 <sup>d</sup>	,839	,789	4,16683	,069	5,557	1	13	,035	1,876

a. Predictores: (Constante), TPACK6

b. Predictores: (Constante), TPACK6, TPK1

c. Predictores: (Constante), TPACK6, TPK1, TCK2

d. Predictores: (Constante), TPACK6, TPK1, TCK2, TPK6

e. Variable dependiente: SumaPK\_CK\_PCK

Fuente: elaboración propia

El análisis ANOVA reflejado en la tabla 6.10 nos muestra que todos los modelos (incluido el modelo 4 que ha sido el elegido), tienen una alta significatividad.

**Tabla 6.10: ANOVA<sup>a</sup>**

Modelo		Suma de cuadrados	df	Media de cuadrados	F	Sig.
1	Regresión	813,342	1	813,342	22,115	,000 <sup>b</sup>
	Residual	588,436	16	36,777		
	Total	1401,778	17			
2	Regresión	956,915	2	478,457	16,133	,000 <sup>c</sup>
	Residual	444,863	15	29,658		
	Total	1401,778	17			
3	Regresión	1079,579	3	359,860	15,636	,000 <sup>d</sup>
	Residual	322,199	14	23,014		
	Total	1401,778	17			
4	Regresión	1176,065	4	294,016	16,934	,000 <sup>e</sup>
	Residual	225,712	13	17,362		
	Total	1401,778	17			

a. Variable dependiente: SumaPK\_CK\_PCK

b. Predictores: (Constante), TPACK6

c. Predictores: (Constante), TPACK6, TPK1

d. Predictores: (Constante), TPACK6, TPK1, TCK2

e. Predictores: (Constante), TPACK6, TPK1, TCK2, TPK6

Fuente: elaboración propia

En la Tabla 6.11 se muestran los coeficientes de los modelos calculados de la manera habitual. Como se puede observar, cada coeficiente representa el efecto adicional de agregar esa variable al modelo, si los efectos de todas las demás variables en el modelo ya son tenidos en cuenta.

**Tabla 6.11 Coeficientes de los modelos generados<sup>a</sup>**

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados Beta	t	Sig.	Estadísticas de colinealidad	
		B	Error estándar				Tolerancia	VIF
1	(Constante)	6,386	11,045		,578	,571		
	TPACK6	12,040	2,560	,762	4,703	,000	1,000	1,000
2	(Constante)	7,603	9,934		,765	,456		
	TPACK6	7,263	3,162	,460	2,297	,036	,529	1,892
	TPK1	4,551	2,069	,440	2,200	,044	,529	1,892
3	(Constante)	13,899	9,166		1,516	,152		
	TPACK6	9,750	2,987	,617	3,265	,006	,460	2,175
	TPK1	6,339	1,980	,613	3,202	,006	,448	2,233
	TCK2	-5,441	2,357	-,424	-2,309	,037	,487	2,052
4	(Constante)	8,770	8,253		1,063	,307		
	TPACK6	8,396	2,657	,531	3,160	,008	,438	2,281
	TPK1	6,087	1,723	,589	3,533	,004	,446	2,242
	TCK2	-7,078	2,162	-,551	-3,274	,006	,437	2,288
	TPK6	4,355	1,847	,338	2,357	,035	,602	1,660

a. Dependent Variable: SumPK\_CK\_PCK

Fuente: elaboración propia

Finalmente, como resultado del método por etapas “Stepwise”, el cuestionario final que se probará en la población final (cuyas variables se volvieron a numerar, para facilitar la interpretación y comprensión del estudio), se presenta en la Tabla 6.12.

**Tabla 6.12: Cuestionario Final**

<b>DIMENSIÓN DE PEDAGOGÍA, CONOCIMIENTO Y CONTENIDO</b>	
<b>ITEM</b>	<b>CUESTIÓN</b>
PK1	Las herramientas utilizadas en la enseñanza, en los cursos relacionados con la alta tecnología computacional, son adecuadas.
PK2	La organización temporal, en los cursos relacionados con la tecnología de la supercomputación, (duración, distribución de los tiempos, etc.) es adecuada.
PK3	El proceso formativo llevado a cabo en los cursos relacionados con la alta tecnología computacional, es adecuado desde el punto de vista pedagógico.
PK4	En los cursos relacionados con la alta tecnología computacional, las clases se desarrollan en un buen ambiente para el aprendizaje.
PK5	El modo de enseñanza es comprensible por parte de los estudiantes, en relación a los contenidos específicos sobre los usos de la tecnología.
PK6	La enseñanza se adapta al nivel de los estudiantes y a los diferentes niveles de conocimiento en el uso de la alta tecnología computacional.
CK1	El contenido de los programas de los cursos relacionados con alta tecnología computacional es adecuado.
CK2	Las explicaciones sobre el uso de la alta tecnología computacional se adaptan a mi especialidad o campo de conocimiento.
CK3	Los métodos de enseñanza-aprendizaje de los cursos relacionados con la alta tecnología computacional son adecuados para una mejor comprensión de los contenidos.
PCK1	El contenido de los cursos relacionados con alta tecnología computacional es práctico.
PCK2	El enfoque de la enseñanza en los cursos relacionados con alta tecnología computacional es adecuado.
PCK3	La evaluación del método de aprendizaje, por parte del alumno, en los cursos relacionados con las altas tecnologías computacionales es adecuada.
PCK4	La formación on-line está adaptada a cursos relacionados con altas tecnologías computacionales.
PCK5	Las altas tecnologías computacionales facilitan la enseñanza.
<b>DIMENSIÓN TECNOLÓGICA</b>	
<b>ITEM</b>	<b>CUESTIÓN</b>
TPK1	El software utilizado en la capacitación que se imparte en los cursos relacionados con la alta tecnología informática es adecuado.
TPK2	El profesor plantea temas y casos prácticos sobre cómo utilizar la alta tecnología computacional.
TCK1	Las altas tecnologías computacionales son una buena herramienta para mejorar la comprensión del objeto de los cursos de formación relacionados con la tecnología.
TPACK1	El uso de altas tecnologías computacionales mejora la confianza para aprender a resolver los problemas de mi especialidad.

Fuente: elaboración propia

En el Apéndice IV se detalla la evolución del constructo desarrollado, analizando los ítems que se han ido eliminando en cada fase hasta llegar al cuestionario final.

#### **6.4.5.- Estudio final**

La encuesta fue remitida a los estudiantes registrados en la Base de Datos del Centro de Supercomputación de Castilla y León (SCAYLE), dentro de los alumnos que asistieron a sesiones formativas organizadas por el Centro. Para su cumplimentación se utilizó *Google Drive Survey*, calculando tanto frecuencias simples como tabulaciones cruzadas. Se utilizó un diseño de encuesta factorial. Finalmente, 97 estudiantes respondieron al cuestionario.

Una vez finalizado el proceso de simplificación del cuestionario (a través de un análisis previo en la prueba piloto), el siguiente paso fue determinar el número óptimo de factores o dimensiones mediante un análisis factorial exploratorio de los componentes principales. Esta es una técnica estadística utilizada para reducir los datos a un conjunto más pequeño de variables de resumen y para explorar la estructura teórica subyacente de los fenómenos, en casos como este, en el campo de la evaluación, donde el número de elementos, en relación con la escala original escala, baja significativamente.

Finalmente, realizamos un análisis factorial en cada dominio de sub-conocimiento de la medida TPACK, para probar nuestros datos, con el fin de confirmar la similitud entre su estructura y la encontrada en el cuestionario TPACK original (Schmidt, et al., 2009). La validez del constructo se estudió para cada subescala de dominio de conocimiento, utilizando el análisis factorial de componentes principales con rotación Varimax y normalización de Kaiser, también comparando los resultados con una rotación Oblimin. Según lo medido por valores propios (mayores que 1) y cargas de factores, estos resultados mostraron un grado relativamente alto de validez de constructo.

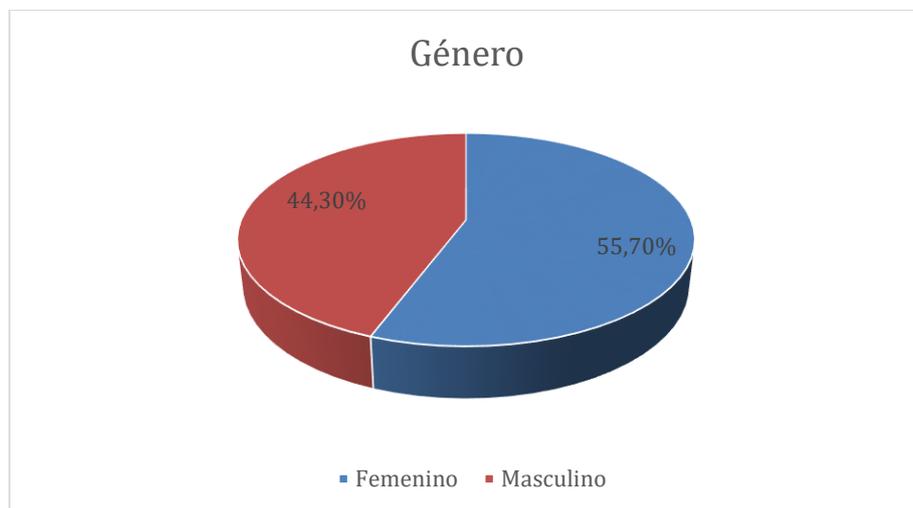
En el siguiente apartado se desarrolla en mayor detalle el resultado final del estudio.

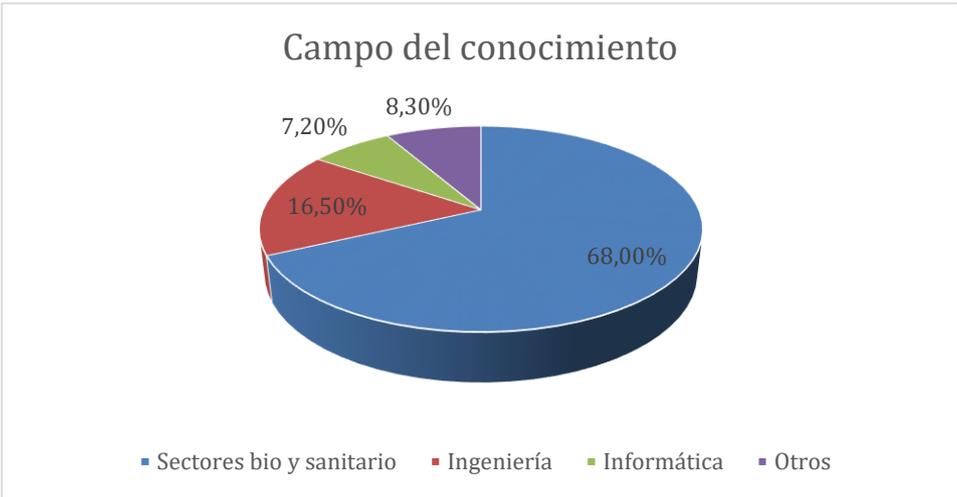
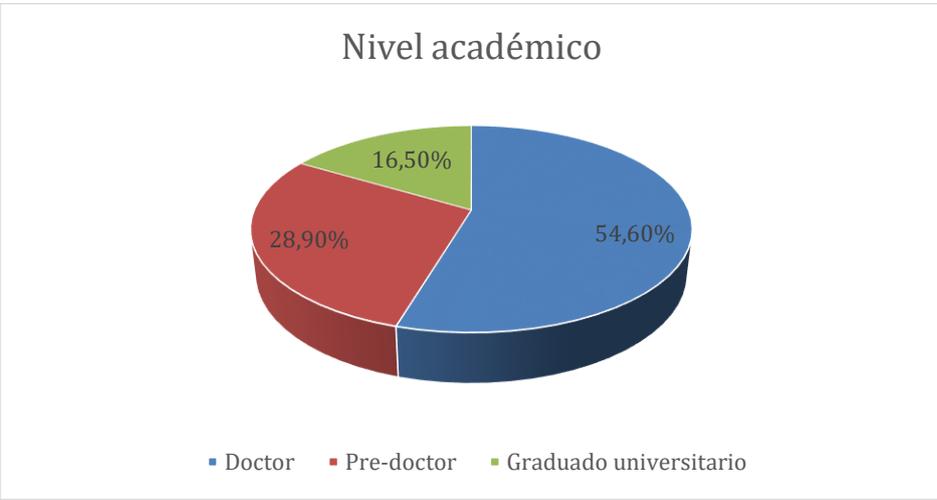
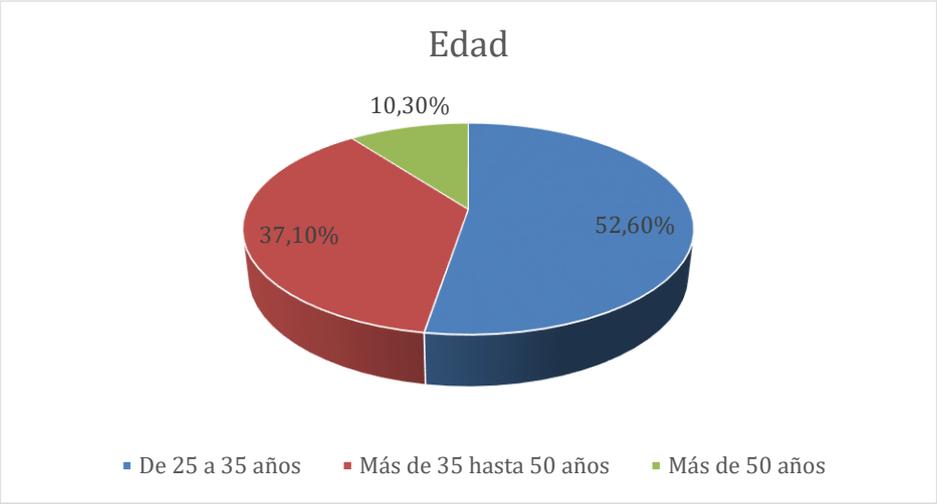
## 6.5.- RESULTADOS

Para llevar a cabo la descripción de los resultados del estudio, debemos considerar el hecho de que era la primera vez que se administraba una adaptación de un modelo TPACK en el campo de la supercomputación. Para obtener los resultados adecuados, hemos evaluado la escala de dimensionalidad y confiabilidad.

La mayoría de los 97 encuestados pertenecían al grupo de edad de 25 a 35 años. Este porcentaje disminuye a partir de los 35 años, lo que indica una población de encuestados relativamente joven. El 55,7% de los encuestados eran mujeres; El 68% de los encuestados pertenecen al sector bio; El 54,6% tenía estudios de nivel de doctorado. El 76.3% consideró que tenían acceso fácil a las instalaciones tecnológicas, pero solo el 45.4% declaró tener acceso fácil a la formación ofrecida por sus organizaciones. Toda la información se presenta en la Figura 6.2.

**Figura 6.2: Representación gráfica de los datos generales de la encuesta**





Fuente: elaboración propia

Una vez que se estudió la información general, se realizó un análisis factorial para agrupar los ítems individuales en un menor número de dimensiones con la finalidad de probar la

validez de constructo del cuestionario (Bornstedt, 1977; Rattray y Jones, 2007) y poder analizar si todos los ítems juntos representan bien la construcción subyacente. El análisis factorial exploratorio detecta los constructos (factores) que subyacen a un conjunto de datos basado en las correlaciones entre variables (ítems del cuestionario) (Field, 2009). Este proceso se emplea para identificar variables o construcciones latentes y se puede usar tanto para simplificar datos, como para reducir el número de variables en los modelos de regresión.

La rotación realizada fue la Varimax que representa la más utilizada según la literatura y se considera la opción más común para analizar la validez de constructo. Para un estudio más amplio, también se estudió la rotación de Oblimin.

En primer lugar, para poder conocer si es recomendable realizar un análisis factorial, se muestran la matriz anti-imagen y el Kaiser-Meyer-Olkin (KMO) y la Prueba de Esfericidad de Bartlett. La matriz anti-imagen de covarianzas y correlaciones es la negativa del coeficiente de correlación parcial. La información sobre la matriz de correlación debe contener pocos valores altos en términos absolutos y no debe haber una gran cantidad de coeficientes cero, de lo contrario, se recomienda no realizar un análisis factorial. La prueba KMO es una medida de cuán adecuados son los datos para el análisis factorial. La prueba mide la adecuación del muestreo para cada variable en el modelo y para el modelo en su conjunto. La estadística es una medida de la proporción de varianza entre las variables que podrían ser varianza común. Los valores altos (cerca de 1.0) generalmente indican que un análisis factorial puede ser útil. La Prueba de Esfericidad de Bartlett se utiliza para probar la hipótesis nula, verificando que la matriz de correlación es una matriz de identidad, lo que indica que las variables no están relacionadas y, por lo tanto, no son adecuadas para la detección de estructuras. Los valores pequeños (menos de 0.05) del nivel de significancia indican que un análisis factorial puede ser útil con sus datos.

Los resultados de estas pruebas que nos permitirán conocer la pertinencia de realizar un análisis factorial son los siguientes:

**Tabla 6.13: Matrices Anti-imagen**

	PK1	PK2	PK3	PK4	PK5	PK6	CK1	CK2	CK3	PCK1	PCK2	PCK3	PCK4	PCK5	TPK1	TPK2	TCK1	TPACK1	
Matriz Anti-imagen de Covarianzas	PK1	,323	-,072	,006	-,009	-,013	-,033	,028	,003	-,073	,020	-,004	,017	,003	-,022	-,043	-,049	,010	,003
	PK2	-,072	,346	-,057	,015	,027	-,062	-,022	-,006	,025	-,015	-,021	-,028	-,067	-,025	-,079	,025	,066	-,007
	PK3	,006	-,057	,194	-,032	-,027	-,001	-,036	,041	-,022	,033	-,066	-,009	,054	,032	-,008	-,038	,004	,012
	PK4	-,009	,015	-,032	,295	-,128	,033	-,043	,028	-,031	-,012	-,009	,047	,018	-,016	-,008	-,032	,008	-,011
	PK5	-,013	,027	-,027	-,128	,264	-,076	,016	-,010	,023	-,039	-,021	-,011	-,025	,035	-,060	,063	,017	-,084
	PK6	-,033	-,062	-,001	,033	-,076	,258	-,040	-,007	-,027	-,005	-,010	-,026	,040	,026	-,004	,002	-,073	,041
	CK1	,028	-,022	-,036	-,043	,016	-,040	,220	-,049	-,028	-,087	,029	-,023	-,026	-,022	,013	,009	-,062	,018
	CK2	,003	-,006	,041	,028	-,010	-,007	-,049	,277	-,090	-,014	,000	,001	,003	-,034	-,100	,001	,069	-,061
	CK3	-,073	,025	-,022	-,031	,023	-,027	-,028	-,090	,176	,008	-,018	-,056	-,009	,038	,033	-,001	-,024	,019
	PCK1	,020	-,015	,033	-,012	-,039	-,005	-,087	-,014	,008	,273	-,008	-,023	,006	,025	-,016	-,083	,043	-,009
	PCK2	-,004	-,021	-,066	-,009	-,021	-,010	,029	,000	-,018	-,008	,149	-,047	-,048	-,039	,003	-,037	-,016	,022
	PCK3	,017	-,028	-,009	,047	-,011	-,026	-,023	,001	-,056	-,023	-,047	,217	,045	-,029	,010	,003	,008	-,028
	PCK4	,003	-,067	,054	,018	-,025	,040	-,026	,003	-,009	,006	-,048	,045	,688	-,128	-,071	,077	-,031	,007
	PCK5	-,022	-,025	,032	-,016	,035	,026	-,022	-,034	,038	,025	-,039	-,029	-,128	,278	,068	-,011	-,108	-,120
	TPK1	-,043	-,079	-,008	-,008	-,060	-,004	,013	-,100	,033	-,016	,003	,010	-,071	,068	,384	-,017	-,061	-,020
	TPK2	-,049	,025	-,038	-,032	,063	,002	,009	,001	-,001	-,083	-,037	,003	,077	-,011	-,017	,174	-,045	-,056
TCK1	,010	,066	,004	,008	,017	-,073	-,062	,069	-,024	,043	-,016	,008	-,031	-,108	-,061	-,045	,349	-,012	
TPACK1	,003	-,007	,012	-,011	-,084	,041	,018	-,061	,019	-,009	,022	-,028	,007	-,120	-,020	-,056	-,012	,257	
Matriz Anti-imagen de Correlaciones	PK1	<b>,963<sup>a</sup></b>	-,215	,025	-,030	-,046	-,114	,106	,011	-,307	,066	-,019	,066	,006	-,073	-,123	-,206	,029	,012
	PK2	-,215	<b>,949<sup>a</sup></b>	-,222	,048	,091	-,206	-,078	-,020	,102	-,047	-,092	-,101	-,137	-,080	-,216	,101	,191	-,024
	PK3	,025	-,222	<b>,943<sup>a</sup></b>	-,135	-,120	-,002	-,173	,178	-,122	,144	-,387	-,043	,149	,139	-,029	-,207	,016	,052
	PK4	-,030	,048	-,135	<b>,942<sup>a</sup></b>	-,458	,120	-,167	,097	-,137	-,042	-,045	,186	,040	-,057	-,023	-,142	,026	-,041
	PK5	-,046	,091	-,120	-,458	<b>,902<sup>a</sup></b>	-,292	,065	-,039	,107	-,144	-,106	-,046	-,058	,129	-,190	,296	,057	-,321
	PK6	-,114	-,206	-,002	,120	-,292	<b>,954<sup>a</sup></b>	-,169	-,025	-,125	-,019	-,051	-,109	,096	,099	-,014	,011	-,244	,157
	CK1	,106	-,078	-,173	-,167	,065	-,169	<b>,947<sup>a</sup></b>	-,197	-,143	-,354	,160	-,106	-,066	-,089	,045	,047	-,225	,076
	CK2	,011	-,020	,178	,097	-,039	-,025	-,197	<b>,928<sup>a</sup></b>	-,406	-,050	,001	,005	,007	-,122	-,307	,006	,223	-,230
	CK3	-,307	,102	-,122	-,137	,107	-,125	-,143	-,406	<b>,935<sup>a</sup></b>	,038	-,112	-,289	-,026	,170	,126	-,004	-,095	,090
	PCK1	,066	-,047	,144	-,042	-,144	-,019	-,354	-,050	,038	<b>,946<sup>a</sup></b>	-,040	-,093	,015	,092	-,048	-,382	,141	-,034
	PCK2	-,019	-,092	-,387	-,045	-,106	-,051	,160	,001	-,112	-,040	<b>,951<sup>a</sup></b>	-,261	-,151	-,190	,012	-,231	-,068	,115
	PCK3	,066	-,101	-,043	,186	-,046	-,109	-,106	,005	-,289	-,093	-,261	<b>,963<sup>a</sup></b>	,116	-,117	,034	,013	,029	-,119
	PCK4	,006	-,137	,149	,040	-,058	,096	-,066	,007	-,026	,015	-,151	,116	<b>,575<sup>a</sup></b>	-,293	-,138	,221	-,063	,017
	PCK5	-,073	-,080	,139	-,057	,129	,099	-,089	-,122	,170	,092	-,190	-,117	-,293	<b>,864<sup>a</sup></b>	,210	-,048	-,348	-,447
	TPK1	-,123	-,216	-,029	-,023	-,190	-,014	,045	-,307	,126	-,048	,012	,034	-,138	,210	<b>,943<sup>a</sup></b>	-,064	-,168	-,062
	TPK2	-,206	,101	-,207	-,142	,296	,011	,047	,006	-,004	-,382	-,231	,013	,221	-,048	-,064	<b>,928<sup>a</sup></b>	-,183	-,267
TCK1	,029	,191	,016	,026	,057	-,244	-,225	,223	-,095	,141	-,068	,029	-,063	-,348	-,168	-,183	<b>,920<sup>a</sup></b>	-,039	
TPACK1	,012	-,024	,052	-,041	-,321	,157	,076	-,230	,090	-,034	,115	-,119	,017	-,447	-,062	-,267	-,039	<b>,914<sup>a</sup></b>	

a. Medidas de adecuación de la muestra (MAM) Fuente: elaboración propia

**Tabla 6.14: Test KMO y Bartlett**

Medida Kaiser-Meyer-Olkin de adecuación de la muestra.		,935
Test de esfericidad de Bartlett	Approx. Chi <sup>2</sup>	1560,498
	Grados de libertad	153
	Sig.	,000

Fuente: elaboración propia

En relación con la idoneidad del uso del análisis factorial, las dos tablas mencionadas indican: (1) la diagonal de las correlaciones los valores están cerca de 1 en la matriz anti-imagen (Tabla 6.13), lo que indica que es apropiado hacer un análisis factorial; (2) el índice KMO (Tabla 6.14) es mayor que 0.5, para la mayoría cercana a 1 ( $> 0.7$ ) indicando (a), que es apropiado emplear el análisis factorial, y (b), la significancia del índice de Bartlett es alta ( $p < 0.05$ ), rechazando la hipótesis nula y también indicando que continuar con el análisis es adecuado.

Para obtener los resultados del estudio, primero se efectuará un análisis factorial y posteriormente se llevará a cabo una rotación para obtener los resultados definitivos.

### **6.5.1.- Análisis factorial**

Una vez que se ha determinado que el análisis factorial es adecuado, procedemos a determinar los parámetros principales. Los pasos son los siguientes: (1) análisis de las comunalidades; (2) estudio de la varianza total explicada; (3) matriz de componentes y finalmente rotación.

Las comunalidades, representadas en la Tabla 6.15 indican la varianza común compartida por factores con variables dadas. Una comunalidad más alta indica que la solución del factor ha extraído una mayor cantidad de la varianza en la variable. Para una mejor medición del análisis factorial, es aconsejable eliminar cualquier elemento con un puntaje de comunalidad de menos de 0.2 (Child, 2006). En este caso, las comunalidades son altas,

lo que significa que una gran proporción de la varianza se explica por factores comunes, no existiendo elementos susceptibles de eliminar.

**Tabla 6.15: Comunalidades**

	Inicial	Extracción
PK1	1,000	,660
PK2	1,000	,596
PK3	1,000	,779
PK4	1,000	,625
PK5	1,000	,593
PK6	1,000	,727
CK1	1,000	,742
CK2	1,000	,627
CK3	1,000	,780
PCK1	1,000	,681
PCK2	1,000	,813
PCK3	1,000	,756
PCK4	1,000	,637
PCK5	1,000	,787
TPK1	1,000	,545
TPK2	1,000	,748
TCK1	1,000	,584
TPACK1	1,000	,697

Extraction Method: Principal Component Analysis.

Fuente: elaboración propia

La varianza total explicada (Tabla 6.16) muestra cómo se divide la varianza entre los 18 posibles factores, en los que analizamos, según el criterio establecido, los dos primeros componentes, con un valor propio mayor que 1. La varianza explicada por estos dos factores es 68.764%, lo cual es un valor aceptable, según algunos autores (Beavers et al., 2013).

**Tabla 6.16: Varianza total explicada**

Componente	Valores propios iniciales			Sumas de extracción de cargas cuadradas			Sumas de rotación de cargas cuadradas <sup>a</sup>
	Total	% de Varianza	% Acum.	Total	% de Varianza	% Acum.	Total
1	10,928	60,709	60,709	10,928	60,709	60,709	10,855
2	1,450	8,054	68,764	1,450	8,054	68,764	2,665
3	,890	4,946	73,709				
4	,757	4,204	77,913				
5	,609	3,383	81,296				
6	,502	2,787	84,082				
7	,456	2,535	86,617				
8	,403	2,236	88,854				
9	,380	2,111	90,965				
10	,308	1,711	92,676				
11	,281	1,560	94,236				
12	,194	1,077	95,313				
13	,183	1,016	96,329				
14	,170	,944	97,273				
15	,162	,902	98,175				
16	,120	,667	98,842				
17	,106	,588	99,430				
18	,103	,570	100,000				

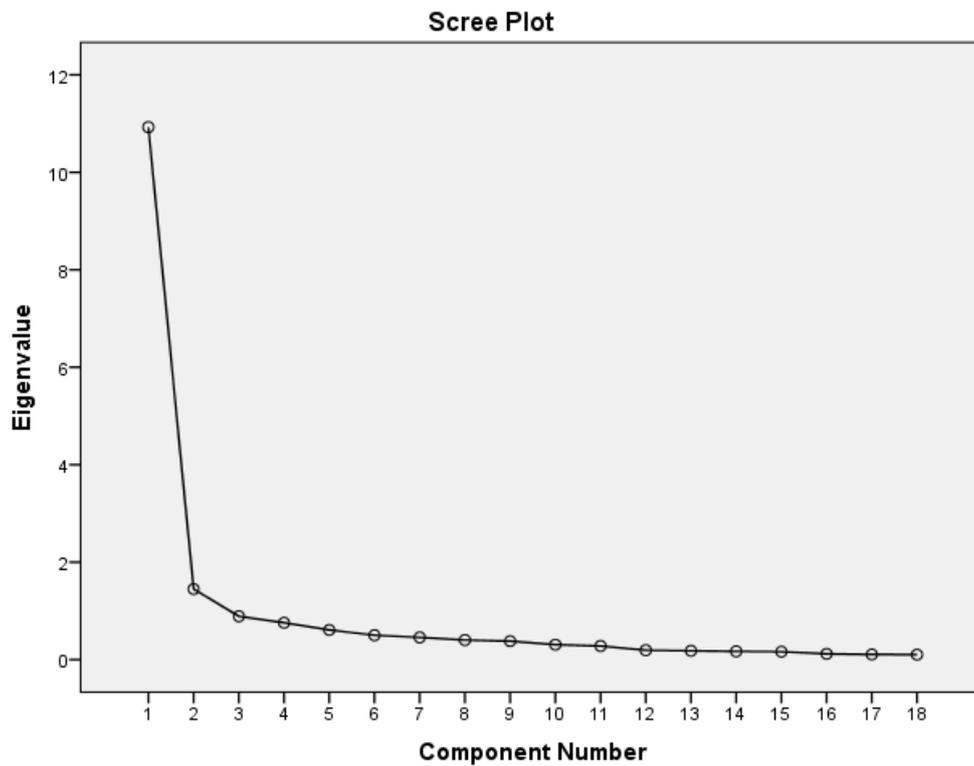
Método de extracción: Análisis de componentes principales

a. Cuando los componentes están correlacionados, no se pueden sumar cargas cuadradas para obtener una varianza total.

Fuente: elaboración propia

El diagrama de pantalla (Gráfico 6.1) representa la información de la varianza total explicada en un gráfico de los factores y sus valores propios correspondientes. En el eje X, los factores (componentes) están representados y los valores propios (se explica la varianza total) están a lo largo del eje Y. El número de componente puede confirmar la validez de los dos primeros componentes, lo que explica el 68,76% de la varianza.

**Gráfico 6.1: Gráfico de sedimentación**



Fuente: elaboración propia

Finalmente, observamos la matriz de componentes (Tabla 6.17), que muestra las cargas de factores basadas en los coeficientes de correlación entre las variables (filas) y los factores (columnas). Cuanto mayor sea el valor absoluto de la carga del factor, más fuerte será la conexión entre la variable y el factor. Un problema común al interpretar la matriz de cargas de factores no rotados es que todas las cargas más significativas se concentran en uno o dos de los primeros factores.

**Tabla 6.17: Matriz de Componentes<sup>a</sup>**

	Componente	
	1	2
PK1	,807	
PK2	,770	
PK3	,839	
PK4	,785	
PK5	,767	
PK6	,823	
CK1	,861	
CK2	,787	
CK3	,865	
PCK1	,819	
PCK2	,901	
PCK3	,868	
PCK4		,788
PCK5	,625	,629
TPK1	,738	
TPK2	,864	
TCK1	,711	
TPACK1	,740	

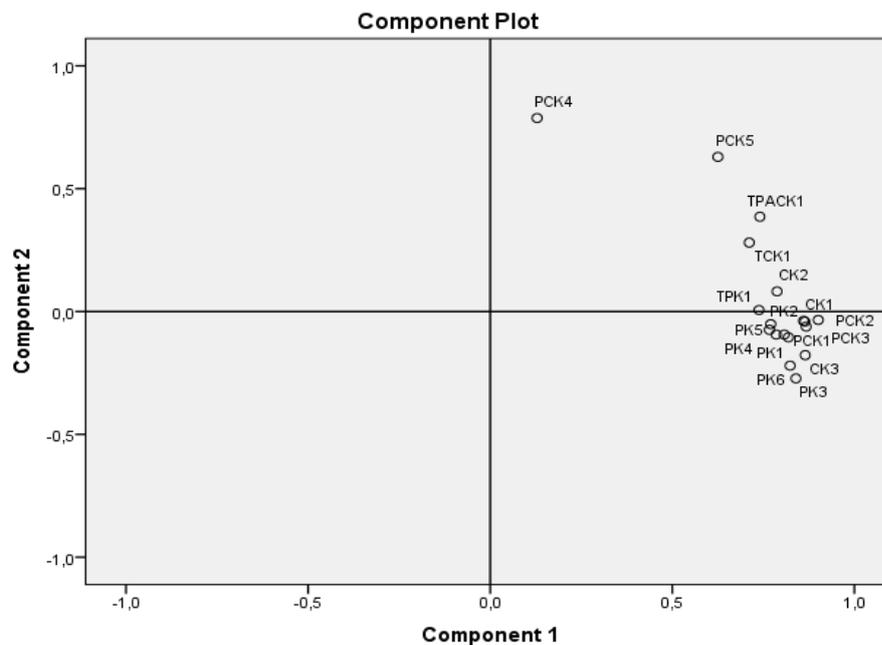
Método de extracción: Análisis de componentes principales.

a. 2 componentes extraídos.

Fuente: elaboración propia

En el gráfico de componentes sin rotación (Gráfico 6.2) se observa que los datos siguen una distribución normal y no se perciben valores atípicos. Como se ve en la matriz de componentes, las variables que se aproximan a los valores - 1 y 1 de cada componente indican un mayor poder de influencia. Por lo tanto, en el componente 1 se observa que PCK2, PCK3, CK1, CK3 y TPK2 (mayores a 0,85) tienen el mayor poder de influencia y en el componente 2, las variables que más influyen son PCK4 y PCK5.

**Gráfico 6.2: Gráfico de componentes no rotados**



Fuente: elaboración propia

### 6.5.2.- Rotación

La rotación se lleva a cabo debido a que la estructura de los factores no es única y, por lo tanto, ciertos grupos de factores que se estimaron originalmente pueden modificarse y dar como resultado un nuevo grupo de factores que funcionan mejor para ciertos problemas particulares. Cuando se extraen varios factores, la interpretación de lo que representan debe basarse en los elementos que se cargan en ellos (Field, 2009). Si diversas variables se cargan en varios factores, será bastante difícil determinar la construcción que representan. Una forma de obtener resultados más interpretables es rotar la solución. Por lo tanto, en el análisis factorial, los factores se rotan hacia algunas variables y lejos de otras. La rotación de factores se acepta generalmente como un paso secuencial cuando se realiza un análisis factorial.

Los estudios de factores rotados se basan en los trabajos de Thurstone (1947) y por Cattell y Horn (1978), quienes defendieron su uso para simplificar la estructura de factores,

haciendo que su interpretación sea más fácil y confiable (es decir, más fácil de replicar con diferentes muestras de datos).

Los factores o componentes y las cargas factoriales de cada variable son combinaciones lineales de estas relaciones. El propósito matemático del análisis factorial es resumir las relaciones entre variables y factores. Una vez que se ha determinado el número de factores a incluir, se descartan todos los demás factores. Los elementos se vuelven a factorizar y, por lo tanto, se obtiene un nuevo número específico de factores. Esa solución se rota en lo que se llama rotación de factores.

Para llevar a cabo rotaciones factoriales, existen principalmente dos métodos: a) el método de rotaciones oblicuas (se aplica a factores que están correlacionados) y b) el método de rotaciones ortogonales (se aplica a factores que no están correlacionados). Estos métodos se basan en el principio de estructura simple y en ninguno de ellos la rotación afecta la bondad del ajuste de la solución factorial, porque incluso si la matriz factorial cambia, las communalidades permanecen sin cambios. Sin embargo, la varianza explicada por cada factor cambia porque depende del método seleccionado.

En este estudio, primero analizaremos la rotación de Oblimin (de tipo oblicuo) para analizar tanto el resultado del análisis factorial como la correlación entre los factores. Si la correlación entre los componentes no es importante, se estimará una rotación Varimax (rotación ortogonal) para comparar los resultados y tener un conocimiento más profundo del fenómeno estudiado.

#### **6.5.2.1.- Rotación Oblimin**

La rotación Oblimin se utiliza para obtener rotaciones oblicuas que permitan transformar vectores, asociados con el análisis de componentes principales o análisis factorial, en estructuras simples. Según los expertos (Thompson, 2004), los coeficientes de una matriz de patrones (Tabla 6.18) son los más importantes en este tipo de rotación y pueden interpretarse de manera similar a un coeficiente de regresión parcial estandarizado, lo que

indica la contribución única de un factor a un elemento mientras controla las contribuciones / efectos de otros factores en ese elemento. De manera similar, cuantifica cuánto aumentaría la valoración para ese elemento (en desviaciones estándar) si ese valor de factor latente aumenta la desviación estándar, mientras mantiene constantes los valores para otros factores. En el caso del coeficiente de la matriz de estructura (Tabla 6.19), se representa la fuerza de la relación entre el indicador / elemento y el factor, al tiempo que ignora la relación de ese factor con todos los demás factores.

**Tabla 6.18:Matriz de patrones<sup>a</sup>**

	Componentes	
	1	2
PK1	,819	
PK2	,767	
PK3	,917	
PK4	,798	
PK5	,772	
PK6	,882	
CK1	,850	
CK2	,734	
CK3	,906	
PCK1	,834	
PCK2	,887	
PCK3	,865	
PCK4		,828
PCK5		,709
TPK1	,714	
TPK2	,854	
TCK1	,586	
TPACK1	,575	,467

Método de extracción: Análisis de Componentes Principales.

Método de rotación: Oblimin con normalización Kaiser.

a. Rotación convergente en 4 iteraciones.

Fuente: elaboración propia

**Tabla 6.19: Matriz de estructura**

	Componentes	
	1	2
PK1	,812	
PK2	,772	
PK3	,860	
PK4	,790	
PK5	,770	
PK6	,840	
CK1	,861	
CK2	,777	
CK3	,877	
PCK1	,825	
PCK2	,900	
PCK3	,870	
PCK4		,781
PCK5	,568	,811
TPK1	,735	
TPK2	,864	
TCK1	,684	,516
TPACK1	,704	,625

Método de extracción: Análisis de Componentes Principales.

Método de rotación: Oblimin con normalización Kaiser.

Fuente: elaboración propia

**Tabla 6.20: Componentes de la matriz de correlación**

Componentes	1	2
1	1,000	,275
2	,275	1,000

Método de extracción: Análisis de Componentes Principales.

Método de rotación: Oblimin con normalización Kaiser.

Fuente: elaboración propia

Los resultados anteriores confirman que las correlaciones entre los componentes (Tabla 6.20) son muy bajas. Esto apunta al hecho de que el modelo de rotación oblicua no es el método óptimo para estos datos, por lo que confirma que la rotación Varimax sería el método de rotación más adecuado. Además, observamos que la matriz de correlación de componentes muestra relaciones muy débiles entre los factores, lo que refuerza la conclusión de que una solución Varimax es mejor que una Oblimin.

#### **6.5.2.2.- Rotación varimax**

El método de rotación Varimax busca maximizar las ponderaciones a nivel de factor, es decir, se espera que cada elemento o variable sea representativo en solo uno de ellos, a fin de minimizar al máximo el número de variables dentro de cada factor. Bajo la rotación ortogonal, la estructura y los coeficientes del patrón son los mismos. Los factores se mantienen sin correlación en la rotación ortogonal, por lo tanto, cuando se investiga el efecto de un factor en un elemento, no hay necesidad de controlar los posibles efectos de otros factores en ese elemento. La tabla 6.21 muestra los coeficientes de la matriz una vez realizada la rotación.

**Tabla 6.21: Matriz de componentes rotados<sup>a</sup>**

	Componentes	
	1	2
PK1	,793	
PK2	,744	
PK3	,882	
PK4	,772	
PK5	,748	
PK6	,850	
CK1	,824	
CK2	,716	
CK3	,875	
PCK1	,807	
PCK2	,861	
PCK3	,839	
PCK4		,786
PCK5		,801
TPK1	,694	
TPK2	,829	
TCK1	,578	,500
TPACK1	,571	,610

Método de extracción: Análisis de Componentes Principales.

Método de rotación: Varimax con normalización Kaiser.

a. Rotación convergente en 4 iteraciones.

Fuente: elaboración propia

Los factores se consideran bien definidos porque al menos tres variables tienen su mayor peso (Kim y Mueller, 1978). En nuestros casos, como se observa en la Matriz de componentes rotados (Tabla 6.21) que un número considerable de variables tienen el mayor peso en Factor 1 y PCK4, PCK5 y TPACK1 en Factor 2. Los pesos o cargas de los indicadores que definen el factor se interpretan como el "coeficientes de correlación" de cada indicador con cada factor. Los pesos nos dicen qué tiene en común cada indicador con todos los indicadores, es decir, pueden interpretarse de manera similar a la correlación total del ítem.

La Matriz de transformación de componentes (Tabla 6.22) muestra las correlaciones utilizadas en la transformación de la Matriz de componentes en la Matriz de componentes rotados. Esta tabla no es importante en la interpretación de los resultados.

**Tabla 6.22: Component Transformation Matrix**

Component	1	2
1	,943	,332
2	-,332	,943

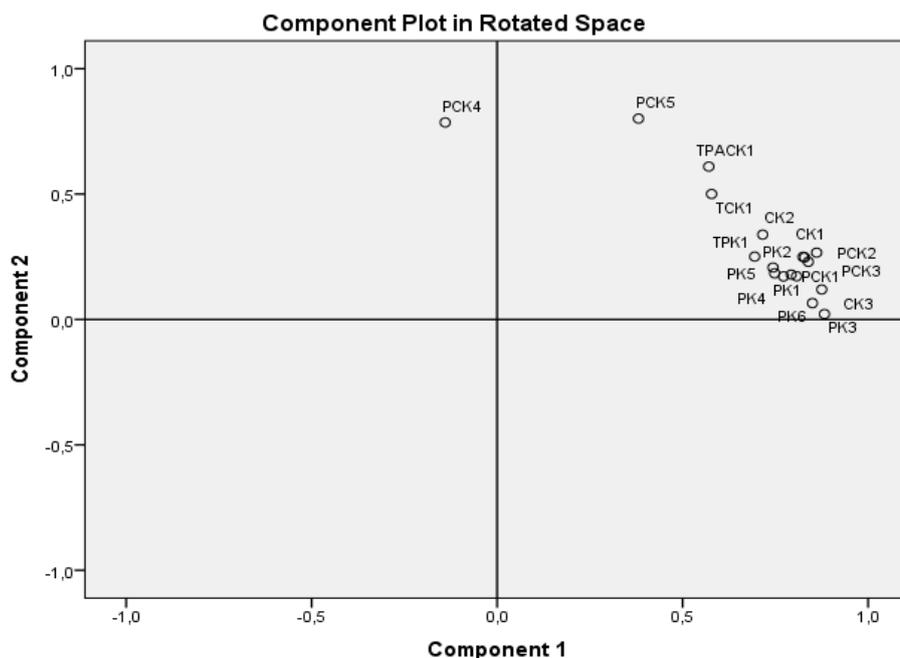
Método de extracción: Análisis de Componentes Principales.

Método de rotación: Varimax con normalización Kaiser.

Fuente: elaboración propia

El gráfico de componentes en el espacio rotado (Gráfico 6.3), utilizando la rotación Varimax, describe la gráfica de puntuación (componente 1 vs. componente 2) y muestra una buena distribución de variables, mostrando que la influencia más importante de cada componente es: en el componente 1, PCK2, PK3, PK6 y CK3 (con valores mayores a 0,85) y en el componente 2, PCK4, PCK5 y TPACK1.

**Gráfico 6.3: Componentes del espacio rotado**



Fuente: elaboración propia

Como se mencionó anteriormente, en términos de cargas más altas, cada factor está vinculado a variables que tienen un significado similar, de esta forma los dos componentes tienen la mayor vinculación con las siguientes variables:

Componente 1:

- PCK2: enfoque de enseñanza adecuado
- PK3: proceso formativo adecuado
- PK6: nivel de alumnos
- CK3: método de enseñanza-aprendizaje adecuado

Componente 2:

- PCK4: entrenamiento en línea
- PCK5: la supercomputación facilita la enseñanza
- TPACK1: el uso de la Supercomputación mejora la confianza.

Por lo tanto, el Componente 1 está más relacionado con los aspectos metodológicos del proceso de enseñanza-aprendizaje, mientras que el Componente 2 está más relacionado con aspectos tecnológicos, relacionados con la influencia de las infraestructuras de supercomputación en la educación. Estos resultados son consistentes, tanto con el modelo TPACK original como con los resultados obtenidos en el presente estudio.

## 6.6.- DISCUSIÓN

Este estudio proporciona un análisis sobre la opinión de grupos heterogéneos de estudiantes en procesos de formación mediante el uso de la supercomputación. Los resultados se basan en un análisis de una encuesta realizada a 97 alumnos de la base de datos del Centro de Supercomputación de Castilla y León (SCAYLE).

Antes de discutir los resultados, es necesario señalar los objetivos generales que motivaron el presente estudio, basado en una muestra de estudiantes que asistieron a cursos relacionados con la supercomputación, que fueron los siguientes: (a) proporcionar información sobre la estructura de factores y las propiedades de una adaptación del modelo TPACK a un cuestionario para este tipo especial de formación y (b) examinar la consistencia y relevancia de los ítems formulados en la adaptación mencionada en el cuestionario para probar la validez de esta forma reducida de la encuesta original TPACK, compuesta exclusivamente por los 18 ítems.

Nuestros hallazgos sugieren que es adecuada una adaptación del cuestionario TPACK original para analizar la opinión de los estudiantes sobre los cursos que utilizan la supercomputación en los procesos de aprendizaje, basado en una reducción de hasta 18 ítems. La obtención de un cuestionario mediante un proceso de diferentes fases, desde la aplicación de la opinión de los expertos hasta la reducción de variables utilizando un modelo Stepwise, con un análisis factorial final, proporciona una excelente explicación de cómo la pedagogía mejora el uso de la tecnología.

El examen de los ítems del cuestionario final muestra que hay dos cuestiones diferentes importantes a la hora de evaluar el uso de la tecnología en la educación. El primero de ellos es la importancia de la pedagogía, el contenido y el conocimiento en el proceso. En segundo lugar, existe la necesidad de complementar los conceptos mencionados anteriormente con habilidades relevantes para la implementación de la tecnología. La relación entre estos diferentes problemas es lógica, ya que, si una persona considera que no tiene ninguna de

las habilidades para implementar la tecnología en la educación, es muy probable que no la utilice.

### **6.6.1.- Resultados**

Actualmente, los resultados obtenidos en el campo de la educación, tienen implicaciones en la implementación de acciones formativas en el campo de la supercomputación. En el presente estudio, se ha podido analizar que el cuestionario final obtenido como resultado de la investigación tiene una consistencia interna muy alta, medida por el alfa de Cronbach (0.955). Esta coherencia no aumentó en exceso al eliminar ninguno de los elementos, por lo que se puede asegurar que la clasificación de los resultados ha sido rápida y simple. Los ítems fueron elegidos con la finalidad de intentar mostrar las mejores implicaciones de la pedagogía, el contenido y el conocimiento en la educación relacionada con la tecnología supercomputacional. Los resultados obtenidos a través de la encuesta desarrollada son alentadores y, a partir del análisis de los mismos, observamos que son consistentes con trabajos anteriores realizados (Schmidt et al., 2009).

Los resultados de la investigación también ofrecen información específica sobre las relaciones entre las dimensiones seleccionadas, observando que los ítems tienen una fuerte correlación, lo que permite deducir que representan un comportamiento instrumental en la mejora de la capacitación. En el análisis factorial realizado, se observó una alta carga de las variables en las dos dimensiones obtenidas en el modelo. Los hallazgos mostraron que los dos factores principales encontrados explican el 68.76% de la varianza, con buena consistencia interna.

Adicionalmente, los resultados sugieren que la escala corta de 18 ítems, que se organiza en dos subescalas basadas en el modelo TPACK, es útil para evaluar la capacitación basada en el uso de la supercomputación. Los resultados del análisis factorial indican que el cuestionario obtenido satisface el criterio de invariabilidad de la estructura interna y puede proporcionar una alternativa favorable para el análisis del modelo TPACK en el caso de  cursos basados en el uso de la supercomputación. El cuestionario demuestra confiabilidad,

consistencia interna y tiene solo dos dimensiones principales en su variación, deduciendo que es conceptualmente único en la medición del fenómeno que se pretende medir: aprovechamiento y calidad de las sesiones de formación basadas en supercomputación. El cuestionario es simple y autoadministrado, requiriendo de unos pocos minutos para ser cumplimentado por el encuestado y unos segundos en ser puntuado por parte del encuestados y ha demostrado ser fácil y rápido de usar y de completar, ofreciendo un gran potencial como herramienta válida y confiable que permita medir la implementación de la supercomputación en el proceso educativo. Con la evidencia obtenida sobre su validez, el cuestionario del estudio podría proporcionar un método muy necesario, simple y estandarizado para medir los resultados de la capacitación a través de la tecnología.

### **6.6.2.- Limitaciones**

Al obtener conclusiones del presente estudio, debemos resaltar cuatro limitaciones principales:

- La primera limitación se refiere a la validez externa. El estudio se basa en una muestra de un solo Centro de Supercomputación, principalmente debido a la existencia de un reducido número de este tipo de centros en España. Para la realización de futuros estudios en línea similar al presente, sería necesario considerar una población más amplia, así como extender los hallazgos actuales a otras muestras más amplias para evaluar la generalización. El mismo argumento se aplicaría a los cursos específicos que se analizaron: relacionados principalmente con la genómica, al ser los más demandados, pero que limita el perfil de los estudiantes al área de biotecnología. Sería de interés analizar un número mayor de estudiantes en otros campos para poder comparar resultados.

- La segunda limitación se refiere a la validez de constructo. El instrumento que utilizamos para evaluar el estudio se basa en una tipología de instrumentos establecidos y validados comúnmente en la literatura, adaptados para medir específicamente el caso del tipo de estudiantes relacionados con formación basada en supercomputación. Este enfoque podría

limitar el número de elementos a analizar necesarios para poder obtener conclusiones más amplias.

- Una tercera limitación es, como se mencionó, que, si bien la validez de constructo y la confiabilidad interna del cuestionario se han demostrado altos en este estudio, sería recomendable realizar un trabajo adicional que permita establecer la sensibilidad de la medida del cambio de actitudes en relación con las actividades educativas. Se requeriría un trabajo adicional, basado en propiedades psicométricas, para poder refinar la medida.

- Una limitación final está relacionada con el análisis factorial realizado. Si bien los componentes principales con rotación Varimax y el criterio de Kaiser son la norma, algunos autores consideran que no son siempre óptimos, particularmente cuando los datos no cumplen con determinados supuestos, como sucede en determinadas ocasiones en el campo de las ciencias sociales. Debido a esto, utilizamos la rotación Oblimin adicionalmente. Hemos visto muchos casos en los que los investigadores utilizaron el Análisis Factorial Exploratorio (EFA) cuando deberían haber usado el análisis factorial confirmatorio. Algunas investigaciones no pueden realizarse fácilmente con EFA, sin embargo, consideramos que la aplicación de estas técnicas ha sido apropiada para nuestro estudio.

## **7.- CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN**

---

### **7.1.- CONCLUSIONES DEL TEMA 3**

Este capítulo proporciona una revisión histórica de las redes de supercomputación y comunicaciones científicas, así como de sus usos actuales y futuros, optimizando el trabajo de las organizaciones, mientras observa que el progreso realizado por la comunidad académica e investigadora ha contribuido históricamente de manera decisiva a los cambios de paradigma. Durante el desarrollo de este estudio, se realizó un análisis en profundidad de la literatura existente, identificando cinco preguntas de investigación para demostrar la importancia de las Redes de Supercomputación y Comunicaciones Científicas en el avance de la ciencia, estableciendo nuevos paradigmas que han permitido y permitirán realizar investigaciones altamente competitivas.

En particular, hemos observado a lo largo de los datos generales recopilados, que la supercomputación ha progresado a gran escala desde su creación en la década de 1940, cuando su uso era en exclusiva para el campo militar, hasta la actualidad, cuando una parte importante de sus usos se destinan más intensamente a la ciencia y en varios campos de conocimiento. Adicionalmente, se ha observado que, cuestiones como la eficiencia energética, son aspectos de gran importancia, lo que plantea un desafío para el futuro por el procesamiento de grandes volúmenes de información, que necesitan grandes redes de comunicación.

Con base en el análisis histórico mencionado anteriormente, podemos destacar las siguientes conclusiones sobre el pasado, el presente y el futuro de los servicios de Supercomputación y las Redes de Comunicaciones Científicas, a través de las respuestas obtenidas por cinco preguntas de investigación exploratoria: (1) la revisión de los principales hitos del pasado permite ofrecer pautas sobre cómo se pueden afrontar los

desafíos del futuro, especialmente con la previsión de la creación en los próximos años de supercomputadoras exaflop; (2) el uso de las supercomputadoras con fines científicos tiene una amplia experiencia práctica, tanto en tiempo de uso como en aplicación a diferentes campos y podemos concluir que casi ninguna investigación científica en el futuro puede regirse sin el uso de herramientas de supercomputación; (3) prácticamente en todos los campos de la ciencia y la industria se experimentará un gran avance mediante el uso de la supercomputación, por lo que los nuevos proyectos y negocios pueden considerar la supercomputación como base sobre la que desarrollar su investigación para crear nuevos productos y/o servicios; (4) el surgimiento de redes de alta velocidad, diferenciadas de la Internet comercial, crea nuevos espacios para compartir, discutir y unir fuerzas sin restricción de espacio, tiempo o distancia para transferir grandes cantidades de datos a través de regiones, países y continentes, por lo que es esencial un desarrollo armonizado de la supercomputación y las redes de comunicación científica; y (5) está claro que, en general, la mayor capacidad de las redes de comunicaciones científicas permite un rendimiento más óptimo de los servicios asociados a la supercomputación.

Este estudio ha concluido que las instalaciones de supercomputación disponibles deben cumplir con el propósito de ser instrumentos adecuados para los procesos de simulación en diversos campos, especialmente cuando cada vez más a menudo la gran mayoría de los problemas deben resolverse mediante un esfuerzo conjunto de múltiples disciplinas científicas. El desarrollo de la investigación colaborativa debe ser la clave que permita optimizar el uso de las supercomputadoras.

Los modelos para simulaciones en supercomputadoras serán algorítmicamente y estructuralmente complejos y contendrán grandes cantidades de información, sobre la base de un gran volumen de datos. Por lo tanto, el hardware debe usarse de manera eficiente, al mismo tiempo que se intenta minimizar el consumo de energía eléctrica. El diseño de redes de interconexión entre procesadores en cada chip y entre nodos del sistema son cuestiones que requieren nuevas ideas, al igual que las redes de comunicación para el intercambio de

datos. En función de lo expuesto, es esencial contar con los medios para capacitar al personal adecuadamente en el uso de estas tecnologías.

Finalmente, debemos tener en cuenta que será necesario utilizar el conocimiento actual relacionado con estos asuntos para proporcionar un punto de partida para futuras investigaciones y explorar nuevos campos donde el uso de la supercomputación será esencial.

## **7.2.- CONCLUSIONES DEL TEMA 4**

Este estudio proporciona dos resultados principales en relación con el estado de la cuestión de la formación en supercomputación: primero, proporciona un resumen de los factores más relevantes para mejorar la formación, así como los factores que mejoran los resultados mediante el uso de una supercomputadora y, segundo, como resultado de la revisión, el análisis de las limitaciones encontradas para un mejor desempeño de los alumnos y las soluciones para estas limitaciones. La intención de este estudio es el de ser una herramienta teórica que ayude en el diseño de actividades y el despliegue de estrategias que, gradualmente, fomenten la mejora de los diferentes componentes de la capacitación mediante el uso de estas infraestructuras. Además, ayuda a crear conciencia sobre el potencial de las supercomputadoras para permitir a los estudiantes un mejor rendimiento.

Finalmente, el trabajo futuro debe centrarse, de manera integral y simultánea, en analizar nuevos métodos y herramientas para evaluar la calidad de la capacitación en supercomputación, y explorar una gama completa de campos donde el uso de las supercomputadoras será útil para mejorar los resultados de las investigaciones. Como se puede observar en el estudio, en el futuro, las sesiones formativas introductorias personalizadas de acuerdo con el conocimiento previo de cada estudiante; la posibilidad de trabajar en línea a través de nuevas redes científicas de comunicaciones; el uso de supercomputadores reales que optimizan los flujos de trabajo o sesiones más prácticas,

deberían ser los factores clave del éxito de los cursos relacionados con la formación en supercomputación.

### **7.3.- CONCLUSIONES DEL TEMA 5**

El documento muestra cómo la implementación de una pedagogía innovadora para la enseñanza de materias relacionadas con STEM para estudiantes que no son STEM y / o no informáticos es muy importante en la sociedad actual. Existe un gran potencial de expansión para los estudios interdisciplinarios y para abordar otras cuestiones importantes a través de las redes y conexiones que creará el personal altamente calificado. Un paso adelante en la integración de la supercomputación en los campos de conocimiento STEM y no STEM a través de un cambio en las políticas podría conducir a una innovación más amplia en la forma tradicional del proceso educativo: mejorar la enseñanza y el aprendizaje.

Actualmente, el uso de herramientas de software para el despliegue de métodos teóricos y análisis de datos utilizados para el modelado matemático, y técnicas de simulación computacional, se extienden ampliamente para el estudio de campos como los sistemas biológicos, a través de la computación de alto rendimiento (Tran y Arabnia, 2015). Esto creará nuevos entornos de aprendizaje que permitirán el mejor uso posible de las tecnologías disponibles, mostrando una forma de mejorar la oferta de sesiones prácticas en los programas de educación superior.

En los cursos relacionados con la supercomputación, los docentes trabajan de una manera orientada a la tecnología basada en situaciones y problemas auténticos que ilustran el potencial de las tecnologías HPC en las prácticas de aprendizaje. Para alcanzar la meta propuesta, la implementación de los cursos debe seguir estos principios: (1) organización de un seminario introductorio que contribuya a un mejor desempeño para el resto del curso; (2) soporte de técnicos especializados e infraestructuras de supercomputación en el sitio, proporcionando un uso óptimo de los recursos computacionales para ejecutar simulaciones,

así como probar la capacidad para resolver problemas prácticos, y (3) instrucciones proporcionadas por expertos altamente calificados.

Podemos concluir que las tecnologías HPC son buenas herramientas pedagógicas, que permiten a los estudiantes aprender de diferentes maneras, en los más diversos campos del conocimiento, considerando que la capacitación en CIS (Ciencias de la Computación e Información) está influenciada en gran medida por la forma en que los científicos hacen uso de big data, el entrenamiento mediante el uso del aprendizaje automático y las técnicas de inteligencia artificial (Yang et al., 2017). Actualmente, el aprendizaje profundo también se describe como una nueva filosofía de capacitación basada en tecnología (Amirian et al., 2019). Al cambiar la forma en que se anuncia y enseña la informática, es posible una informática más inclusiva (Aguar et al., 2016).

En relación con el apoyo brindado por el Centro de Supercomputación en el diseño, implementación y monitoreo de la capacitación ofrecida a los estudiantes, podemos concluir que contribuye a las sesiones de capacitación en los siguientes aspectos positivos:

- Al utilizar las infraestructuras del Centro de Supercomputación para la formación, el trabajo de los estudiantes se controla constantemente. De esta manera, el uso de los recursos se puede optimizar, proporcionando una mayor capacidad computacional si es necesario.
- Se garantiza el soporte continuo de técnicos, expertos en HPC, que cubren varias áreas del uso de la supercomputadora, soluciona problemas y proporciona un alto nivel de seguridad.
- El software involucrado en simulaciones computacionales se está probando permanentemente para ofrecer resultados más confiables.
- Los estudiantes llevan a cabo sus tareas a través del software con la ayuda de técnicos especializados en supercomputación, lo que resulta en un mejor rendimiento.

En la formación se utilizan algoritmos y diagramas de flujo para la identificación de las necesidades de los estudiantes, lo que permite el uso de árboles de decisión de diagnóstico

mediante el uso de una arquitectura distribuida que combina datos compartidos y estilos cliente-servidor (Joshua y Scuse, 2005). Actualmente, se utilizan comúnmente algoritmos que proporcionan un entorno optimizado para resolver problemas de optimización a gran escala, ayudando a los investigadores en la aplicación de algoritmos evolutivos para una amplia gama de campos (Mohammadi et al., 2020). A la luz de esto, el diagrama de flujo descrito en este documento muestra una forma probada de llevar a cabo sesiones de capacitación relacionadas con la supercomputación basadas en la experiencia de los casos estudiados.

Finalmente, el trabajo futuro debe centrarse, de manera integral y simultánea, en analizar nuevos métodos y herramientas para evaluar la calidad de la capacitación en supercomputación con el fin de mejorarla, constantemente adaptada a las necesidades de los estudiantes en cada momento. En consecuencia, es importante explorar los requisitos de todos los campos que el uso de supercomputadoras podría ser útil, tratando de encontrar formas adecuadas de promover dicha capacitación en estudiantes de educación superior.

#### **7.4.- CONCLUSIONES DEL TEMA 6**

El modelo TPACK ha demostrado su efectividad en diferentes niveles educativos, así como en la investigación y la formación docente (Anderson et al., 2013), permitiendo la construcción de modelos teóricos, que mejoran la comprensión del comportamiento de las TIC, que se consideran una herramienta robusta para analizar y reflexionar sobre contextos y procesos de aprendizaje y enseñanza. El estudio de la encuesta desarrollada proporciona una imagen general de los estudiantes que llevan a cabo su formación relacionada con la supercomputación, mostrando cómo el modelo TPACK y sus aplicaciones han ayudado a informar y diseñar pautas e indicadores para futuros programas de formación de desarrollo de tecnología profesional relacionada con la supercomputación.

La validez de constructo del cuestionario evaluado es confiable. Los ítems miden la misma construcción subyacente y la extracción de dos factores parecen ser una consecuencia de la

redacción de los ítems del cuestionario. Los resultados de la encuesta, basada en el modelo TPACK, utilizado en cursos relacionados con la supercomputación, revela que una modificación del cuestionario original a través de un procedimiento orientado al proceso, en primer lugar, por expertos, proporciona una construcción simplificada que ayuda en la medición de la pedagogía, del conocimiento y del contenido basados en el uso de esta tecnología. Hemos observado que un cuestionario más sencillo, basado en 18 ítems, es suficiente para aplicar la adaptación de TPACK a los estudiantes de cursos relacionados con la supercomputación. A través de las opiniones y las puntuaciones de los participantes en la encuesta, hemos podido concluir que una aplicación de este instrumento en el campo de la supercomputación podría ayudar a mejorar las acciones de formación en esta área. Además, la brevedad y la facilidad de aplicación del cuestionario propuesto hacen posible ser eficientes tanto en coste como en velocidad al monitorear este tipo de formación.

Las generaciones más jóvenes de estudiantes se basan más en un conocimiento tecnológico más sofisticado, para lo cual se necesita un adecuado proceso pedagógico para enfrentar los desafíos de las nuevas metodologías. Por lo tanto, podemos concluir que una buena enseñanza requiere una comprensión del conocimiento de cómo la tecnología está relacionada con la pedagogía y el contenido.

El cuestionario sobre formación en supercomputación basado en el modelo TPACK puede ser útil para desarrollar y evaluar programas educativos, desarrollos de políticas e iniciativas de gestión para estudiantes y otros profesionales. Además, puede ser útil para identificar la capacitación tecnológica que más se beneficiaría de las iniciativas educativas. Los resultados de la presente investigación son evidencia a la que se debe agregar en futuros estudios que completen otras facetas de la validez del modelo TPACK. En consecuencia, el trabajo futuro debería incluir la revisión y el refinamiento continuos del instrumento y la ampliación del alcance de la encuesta a otros contextos de educación, incluido el análisis factorial de los datos de una población más amplia de estudiantes para comparar con los resultados de la presente investigación. Además, también se deberían recopilar datos

cualitativos, como entrevistas y observaciones, para proporcionar ilustraciones más detalladas bajo el perfil general del desarrollo del modelo TPACK. Mientras tanto, también se recomienda trabajar en colaboración internacional para proporcionar una visión mundial de TPACK en diferentes contextos culturales.

Este estudio, por lo tanto, alienta y abre un camino para seguir trabajando en el desarrollo de nuevos cuestionarios relacionados con la formación en tecnología y su uso en futuras investigaciones sobre la mejora en la formación en supercomputación. Llegamos a la conclusión de que es necesario continuar investigando para adaptar el modelo TPACK en el campo de la educación tecnológica a fin de mejorar la calidad de los planes de aprendizaje, contribuyendo así al desarrollo de sistemas educativos al ofrecer programas de capacitación adecuados para ayudar a los estudiantes. para lograr sus objetivos.

## 8.- PUBLICACIONES Y DIFUSIÓN DE RESULTADOS

---

Durante el desarrollo de las investigaciones en las que se basó la tesis, se han ido publicando de forma parcial los resultados que se han ido obteniendo, tanto en revistas de impacto, como en congresos relacionados con la temática de estudio.

El detalle de tanto de las publicaciones en revistas de impacto, como en las aportaciones en congresos ha sido la siguiente:

### 8.1.- REVISTAS INDEXADAS

- Fernández, Á., Fernández, C., Miguel-Dávila, J. Á. & Conde, M. Á. (2019). Integrating Supercomputing Clusters into Education: a case study in biotechnology. *The Journal of Supercomputing*. Junio de 2020. DOI 10.1007/s11227-020-03360-5. ISSN 0920-8542.

- Fernández, Á., Fernández, C., Miguel-Dávila, J. Á., Conde, M. Á., & Matellán, V. (2019). Supercomputers to improve the performance in higher education: A review of the literature. *Computers & Education*, 128, 353-364. ISSN: 0360-1315.

- Fernández-González, Á., Rosillo, R., Miguel-Dávila, J. Á., & Matellán, V. (2015). Historical review and future challenges in Supercomputing and Networks of Scientific Communication. *The Journal of Supercomputing*, 71(12), 4476-4503. ISSN 0920-8542.

### 8.2.- CONGRESOS

- Disponibilidad de recursos tecnológicos y necesidades formativas: el caso de la formación en Supercomputación. Fernández, Á., Fernández, C., Miguel-Dávila, J. Á., Conde, M. Á., & Matellán, V. (2019). X Workshop in Operations Management and Technology. 4 y 5 de abril de 2019. Albacete (España).

- Perceptions about availability of technological resources and training needs: a gender perspective based on supercomputing-related courses. Fernández, Á., Fernández, C., Miguel-Dávila, J. Á., Conde, M. Á., & Matellán, V. (2019). 3th International Conference on Industrial Engineering and Industrial Management and XXIII Congreso en Ingeniería en Organización - CIO 2019. 11 y 12 de julio de 2019. Gijón-Asturias (España).
  
- Supercomputers in the educational process. Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality. 16 al 18 de octubre de 2019. León (España).

## APÉNDICE I – Artículos de referencia

---

### Detalle de referencias bibliográficas extraídas utilizando los criterios de calidad.

S1	Abuzagheh, O., Goldschmidt, K., Elleithy, Y., y Lee, J. (2013). Implementing an affordable high-performance computing for teaching-oriented computer science curriculum. <i>ACM Transactions on Computing Education (TOCE)</i> , vol. 13, no. 1.
S2	Armosky, B., Brown, S., Drummond, T., Ferguson, J., Gerber, R., Hacker, T. J., y Traxler, K. (2007). HPC University. In <i>TG08-TeraGrid Conference, Madison, WI</i> .
S3	Bergeron, K., Cummings, R., Decker, C. R., Freeman, M. J., Hoke, C. C., Seidel, J., y MCDaniel, D. M. (2008). An Undergraduate Computational Aerodynamics Curriculum. In <i>DoD HPCMP Users Group Conference, 2008. DOD HPCMP UGC</i> (pp. 18-24). IEEE.
S4	Bernabé, G., Cuenca, J., García, L. P., Giménez, D., y Rivas-Gómez, S. (2014). A High Performance Computing Course Guided by the LU Factorization. <i>Procedia Computer Science</i> , 29, 1446-1457.
S5	Burkhart, H., Guerrero, D., y Maffia, A. (2014). Trusted high-performance computing in the classroom. In <i>Proceedings of the Workshop on Education for High-Performance Computing</i> (pp. 27-33). IEEE Press.
S6	Cuenca, J., y Giménez, D. (2015). TCPP Curriculum in Parallel Programming courses of the Degree in Computer Science at the University of Murcia, Spain. on-line: <a href="http://dis.um.es/~domingo/14/EarlyAdopter/paper.pdf">http://dis.um.es/~domingo/14/EarlyAdopter/paper.pdf</a> .
S7	Czarnul, P. (2014). Teaching high performance computing using beesycluster and relevant usage statistics. In <i>Procedia Computer Science</i> , vol. 29.
S8	Fabricius, U., Freundl, C., Köstler, H., y Rüde, U. (2005). High performance computing education for students in Computational Engineering. In <i>Lecture Notes in Computer Science</i> , vol. 3515, no. II, pp. 27–35.
S9	Farian, H., Anne, K. M., y Haas, M. (2008). Teaching high-performance computing in the undergraduate college CS curriculum. <i>Journal of Computing Sciences in Colleges</i> , 23(3), 135-142.
S10	Fernández Slezak, D., Turjanski, P. G., Montaldo, D., y Mocskos, E. E. (2010). Hands-On Experience in HPC with Secondary School Students. <i>IEEE Transactions on Education</i> , 53(1), 128-135.
S11	Gaziza, B., Salima, N., Guldina, K., y Elena, K. (2012). Parallel Computing in Training of Informatics Teachers. <i>Procedia-Social and Behavioral Sciences</i> , 51, 883-887.
S12	Holmes, V., y Kureshi, I. (2015). Developing High Performance Computing Resources for Teaching Cluster and Grid Computing courses. <i>Procedia Computer Science</i> , 51, 1714-1723.

S13	Liu, X., Guo, H. M., y Meng, C. G. (2009). Design and Implementation of HPC-Based Research-Oriented Learning Environment for Structural Chemistry. In <i>Education Technology and Training, 2009. ETT'09. Second International Conference on</i> (pp. 267-270). IEEE.
S14	Moses, G. A., y Mariasingam, M. A. (2006). The development of a high quality Internet-based degree programme in computational science. 5th Global Congress on Engineering Education, 146-150, Brooklyn, New York, USA, 17-21.
S15	Mullen, J. S., Arcand, W., Bestor, D., Bergeron, W., Byun, C., Gadepally, V., ... y Michaleas, P. (2016) Designing a New High Performance Computing Education Strategy for Professional Scientists and Engineers. <u>IEEE High Performance Extreme Computing Conference (HPEC '16) Twentieth Annual HPEC Conference</u>
S16	Mullen, J., Byun, C., Gadepally, V., Samsi, S., Reuther, A., y Kepner, J. (2017). Learning by doing, High Performance Computing education in the MOOC era. <i>Journal of Parallel and Distributed Computing</i> , 105, 105-115.
S17	Munshi, R., Coalson, R. D., Ermentrout, G. B., Madura, J. D., Meirovitch, H., Stiles, J. R., y Bahar, I. (2006). An introduction to simulation and visualization of biological systems at multiple scales: a summer training program for interdisciplinary research. <i>Biotechnology progress</i> , 22(1), 179-185.
S18	Radenski, A. (2012). Integrating data-intensive cloud computing with multicores and clusters in an HPC course. In <i>Proceedings of the 17th ACM annual conference on Innovation and technology in computer science education - ITiCSE '12</i> .
S19	Ravindran, A. Tolley, P., y Mukherjee, A. (2009). Introducing reconfigurable computing in the undergraduate computer engineering curriculum. <i>Age</i> (Omaha), vol. 14, p. 1, 2009.
S20	Richards, M., and Lathrop, S. (2011). A training roadmap for new HPC users. In <i>Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery</i> (p. 56). ACM.
S21	Ryabinin, K., y Chuprina, S. (2015). Using Scientific Visualization Tools to Bridge the Talent Gap. <i>Procedia Computer Science</i> , 51, 1734-1741.
S22	Shiflet, A. B., y Shiflet, G. W. (2010). Simulating the formation of biofilms in an undergraduate modeling course. <i>Procedia Computer Science</i> , 1(1), 895-901.
S23	Shiflet, A. B., Shiflet, G. W., y Sanders, W. E. (2013). Undergraduate module on computational modeling: Introducing modeling the cane toad invasion. <i>Procedia Computer Science</i> , 18, 1429-1435.
S24	Smith, J. (2014). High Performance Computing and Mathematics. On-line: <a href="http://www.math.iastate.edu/pdfs/MM2014.pdf">http://www.math.iastate.edu/pdfs/MM2014.pdf</a> math.iastate.edu
S25	Smith, J., y Wolfe, G. (2007). Introducing AP computer science students to high-performance computing. <i>Journal of Computing Sciences in Colleges</i> , 23(1), 70-76.

S26	Stainsby, H., Muresano, R., Fialho, L., González, J. C., Rexachs, D., y Luque, E. (2009). Teaching model for computational science and engineering programme. In <i>Computational Science–ICCS 2009</i> (pp. 34-43). Springer Berlin Heidelberg.
S27	Stewart, K. A. (2006). Template for a Group-Work Paradigm in an Undergraduate Supercomputing Course. On-line: <a href="https://www.researchgate.net/publication/2606153_A_Template_for_a_Group-Work_Paradigm_in_an_Undergraduate_Supercomputing_Course">https://www.researchgate.net/publication/2606153_A_Template_for_a_Group-Work_Paradigm_in_an_Undergraduate_Supercomputing_Course</a> .
S28	Stitt, T. y Robinson, T. (2008). A survey on training and education needs for petascale computing. On-line: <a href="http://www.prace-project.eu/IMG/pdf/D3-3-1_document_final.pdf">http://www.prace-project.eu/IMG/pdf/D3-3-1_document_final.pdf</a> .
S29	Stojanovic, N., y Milovanovic, E. (2015). Teaching Introductory Parallel Computing Course with Hands-On Experience. <i>International Journal of Engineering Education</i> , 31(5), 1343-1351.
S30	Touriño, J. Martín, M. J. Tarrío, J., y Arenaz, M. (2005). A grid portal for an undergraduate parallel programming course. <i>Education, IEEE Transactions on</i> , 48(3), 391-399.
S31	Varma, A., Keswani, Y., Bhatnagar, Y., y Chaudhury, B. (2017). Let's HPC: A web-based interactive platform to aid High Performance Computing education. arXiv preprint arXiv:1701.06356. <a href="https://arxiv.org/abs/1701.06356">https://arxiv.org/abs/1701.06356</a>
S32	Vuduc, R., Czechowki, K., Chandramowliswaran, A., y Choi, J. W. (2012). Courses in high-performance computing for scientists and engineers. In <i>Parallel and Distributed Processing Symposium Workshops &amp; PhD Forum (IPDPSW), 2012 IEEE 26th International</i> (pp. 1335-1340). IEEE.
S33	Xiaotao, Y., Aili, L., y Lin, Z. (2010). Research of high performance computing with clouds. In <i>Proc. International Symposium Computer Science and Computational Technology</i> , pp. 289-293.
S34	Yasar, O., Maliekal, J., Little, L. J., y Jones, D. (2006). A computational technology approach to education. <i>Computing in Science and Engineering</i> , 8(3), 76.

## **APÉNDICE II – Encuesta: Versión inicial**

---

Los detalles de los 53 elementos en este primer borrador fueron los siguientes:

### Información personal:

1. Género
2. Edad
3. Campo de conocimiento.
4. nivel académico
5. Institución a la que pertenece

Datos del entorno de trabajo / actividad.

6. ¿Tiene fácil acceso a los recursos tecnológicos que necesita para llevar a cabo su actividad?
7. ¿Su entidad ofrece capacitación adecuada en el uso de tecnologías?

TK

TK1.- Mi nivel en el uso de alta tecnología computacional es adecuado

TK2.- Mis conocimientos previos en el uso de altas tecnologías computacionales son adecuados para mi desempeño profesional.

TK3.- asimilo fácilmente el conocimiento tecnológico

TK4.- Paso el tiempo haciendo pruebas con las herramientas tecnológicas que pueden ayudar a mi actividad profesional.

TK5.- Tengo un conocimiento adecuado de las diferentes tecnologías existentes que sirven para ayudar a mi desempeño profesional.

TK6.- He trabajado con diferentes tipos de herramientas tecnológicas para favorecer mi desempeño profesional.

TK7.- Mi nivel de uso de alta tecnología informática no es adecuado.

PK

PK1.- Las herramientas de enseñanza en cursos relacionados con altas tecnologías computacionales son adecuadas.

PK2.- El tiempo de los cursos relacionados con alta tecnología computacional (duración, distribución temporal, etc.) es adecuado.

PK3.- Existe colaboración o trabajo en equipo con otros grupos / personas en la capacitación impartida en los cursos relacionados con la alta tecnología informática.

PK4.- La pedagogía en cursos de informática de alta tecnología es adecuada.

PK5.- Mejorar el rendimiento del alumno en clase utilizando tecnologías como complemento de la teoría.

PK6.- La clase se maneja adecuadamente en los cursos relacionados con alta tecnología computacional.

PK7.- La forma de enseñar es comprensible para el alumno en relación con los contenidos específicos de los usos de la tecnología.

PK8.- La enseñanza se adapta al nivel de los alumnos y a los diferentes grados de conocimiento de los mismos en el uso de alta tecnología computacional.

PK9.- Hay satisfacción en cursos relacionados con alta tecnología computacional.

PK10.- Se evalúa adecuadamente el grado de conocimiento adquirido en los cursos relacionados con alta tecnología computacional.

PK11.- La tecnología informática de alta tecnología ha mejorado el nivel de los estudiantes.

PK12.- La pedagogía en los cursos de alta tecnología no es adecuada.

## CK

CK1.- El contenido de los programas de los cursos relacionados con la alta tecnología computacional es adecuado.

CK2.- La preparación y calificación del personal docente es adecuada para el curso relacionado con alta tecnología computacional.

CK3.- La explicación sobre el uso de alta tecnología informática se adapta a mi especialidad o campo de conocimiento.

CK4.- El método de enseñanza-aprendizaje de los cursos relacionados con alta tecnología computacional es adecuado.

CK5.- El método de enseñanza-aprendizaje de cursos relacionados con alta tecnología computacional no es adecuado.

## TPK

TPK1.- El software utilizado en la capacitación que se imparte en los cursos relacionados con la alta tecnología informática es adecuado.

TPK2.- Los laboratorios de computación están disponibles para complementar los cursos relacionados con la alta tecnología computacional.

TPK3.- Existen suficientes infraestructuras tecnológicas disponibles para la capacitación en los cursos relacionados con la tecnología informática.

TPK4.- Existe un fácil acceso a la infraestructura de alta tecnología en capacitación.

TPK5.- Hay calidad en los servicios tecnológicos que permiten mejorar la enseñanza.

TPK6.- Las altas tecnologías computacionales permiten mejorar la calidad de la enseñanza.

TPK7.- Las altas tecnologías computacionales son importantes en la formación futura.

TPK8.- El profesor plantea formas y casos prácticos sobre cómo usar alta tecnología computacional.

TPK9.- Las altas tecnologías computacionales son esenciales hoy en día en investigación y desempeño profesional.

TPK10.- No existe calidad en los servicios tecnológicos que permitan mejorar la enseñanza.

TCK

TCK1.- Estoy aprendiendo sobre alta tecnología informática para mejorar mi desempeño profesional (en docencia e investigación).

TCK2.- Las altas tecnologías computacionales mejoran la comprensión del objeto del curso.

TCK3.- Hay dificultades para encontrar un curso relacionado con altas tecnologías computacionales adecuadas para mi desempeño profesional.

TCK4.- Las altas tecnologías computacionales no mejoran la comprensión del objeto del curso.

PCK

PCK1.- El contenido de los cursos relacionados con las altas tecnologías computacionales es práctico.

PCK2.- Existe interés en el contenido de los cursos relacionados con las altas tecnologías computacionales.

PCK3.- El enfoque de enseñanza en cursos relacionados con altas tecnologías computacionales es adecuado.

PCK4.- El método de aprendizaje del alumno en los cursos relacionados con altas tecnologías computacionales es adecuado.

PCK5.- La capacitación en línea se realiza en cursos relacionados con la alta tecnología informática.

PCK6.- Las altas tecnologías computacionales facilitan la enseñanza.

PCK7.- El contenido de los cursos relacionados con las altas tecnologías computacionales no es práctico.

#### TPACK

TPACK1.- El uso de altas tecnologías computacionales en cursos de capacitación mejora la especialización en mi materia.

TPACK2.- La combinación de los contenidos de los cursos relacionados con la alta tecnología computacional, con el soporte tecnológico y la forma en que el profesor imparte el curso es correcto

TPACK3.- Los cursos relacionados con las altas tecnologías computacionales permiten mejorar lo que se enseña y cómo se enseña.

TPACK4.- El material didáctico que se proporciona en los cursos relacionados con las altas tecnologías computacionales es adecuado

TPACK5.- El uso de alta tecnología computacional mejora el contenido de las lecciones.

TPACK6.- El uso de altas tecnologías computacionales mejora el rendimiento en mi desempeño profesional

TPACK7.- La confianza se mejora al aprender el uso de altas tecnologías computacionales para resolver los problemas de mi especialidad.

TPACK8.- El uso de alta tecnología computacional no mejora el contenido de las lecciones.

### APÉNDICE III – Análisis de los elementos de la encuesta

ITEM	MEDICIÓN	
	CVC	IVC
1. Género	0,87	0,6
2. Edad	0,88	0,73
3. Campo del conocimiento	0,81	0,33
4. Nivel académico	0,88	0,67
5. Institución	0,84	0,27
6. ¿Tiene fácil acceso a los recursos tecnológicos que necesita para llevar a cabo su actividad?	0,9	0,8
7. ¿Su entidad ofrece capacitación adecuada en el uso de tecnologías?	0,89	0,93
TK1	0,85	0,53
TK2	0,87	0,67
TK3	0,77	0,47
TK4	0,8	0,47
TK5	0,87	0,73
TK6	0,91	0,73
TK7	0,54	-0,33
PK1	0,9	0,67
PK2	0,84	0,4
PK3	0,85	0,73
PK4	0,89	0,8
PK5	0,91	0,87
PK6	0,77	0,4
PK7	0,91	0,8
PK8	0,93	0,8
PK9	0,71	-0,07
PK10	0,7	0,33
PK11	0,69	0,07
PK12	0,56	-0,33
CK1	0,93	0,93
CK2	0,97	1
CK3	0,93	0,93
CK4	0,95	1

CK5	0,6	-0,4
TPK1	0,93	0,93
TPK2	0,84	0,53
TPK3	0,93	0,8
TPK4	0,83	0,6
TPK5	0,64	-0,27
TPK6	0,76	0,27
TPK7	0,91	0,93
TPK8	0,92	0,87
TPK9	0,92	0,8
TPK10	0,55	-0,27
TCK1	0,86	0,6
TCK2	0,77	0,13
TCK3	0,87	0,6
TCK4	0,55	-0,4
PCK1	0,91	0,73
PCK2	0,77	0,13
PCK3	0,88	0,73
PCK4	0,89	0,87
PCK5	0,81	0,4
PCK6	0,81	0,27
PCK7	0,59	-0,33
TPACK1	0,93	0,87
TPACK2	0,93	0,93
TPACK3	0,75	0,4
TPACK4	0,95	0,93
TPACK5	0,95	0,93
TPACK6	0,87	0,73
TPACK7	0,8	0,4
TPACK8	0,57	-0,27

## APÉNDICE IV – Evaluación experta: Resultados

Item	Comité de expertos	Análisis de confiabilidad y consistencia de la prueba piloto	Cuestionario final
TK1	OK	OK	Eliminado
TK2	OK	OK	Eliminado
TK3	Eliminado		
TK4	OK	OK	Eliminado
TK5	OK	OK	Eliminado
TK6	OK	OK	Eliminado
TK7	Eliminado		
PK1	OK	OK	OK
PK2	OK	OK	OK
PK3	OK	Eliminado	
PK4	OK	OK	OK
PK5	OK	Eliminado	
PK6	OK	OK	OK
PK7	OK	OK	OK
PK8	OK	OK	OK
PK9	Eliminado		
PK10	Eliminado		
PK11	Eliminado		
PK12	Eliminado		
CK1	OK	OK	OK
CK2	OK	Eliminado	
CK3	OK	OK	OK
CK4	OK	OK	OK
CK5	Eliminado		
TPK1	OK	OK	OK
TPK2	OK	OK	Eliminado
TPK3	OK	OK	Eliminado
TPK4	OK	OK	Eliminado
TPK5	Eliminado		

TPK6	Eliminado		
TPK7	OK	OK	Eliminado
TPK8	OK	OK	OK
TPK9	OK	OK	Eliminado
TPK10	Eliminado		
TCK1	OK	OK	Eliminado
TCK2	OK	OK	OK
TCK3	OK	Eliminado	
TCK4	Eliminado		
PCK1	OK	OK	OK
PCK2	Eliminado		
PCK3	OK	OK	OK
PCK4	OK	OK	OK
PCK5	OK	OK	OK
PCK6	OK	OK	OK
PCK7	Eliminado		
TPACK1	OK	Eliminado	
TPACK2	OK	Eliminado	
TPACK3	Eliminado		
TPACK4	OK	OK	Eliminado
TPACK5	OK	OK	Eliminado
TPACK6	OK	OK	Eliminado
TPACK7	OK	OK	OK
TPACK8	Eliminado		

## BIBLIOGRAFÍA

---

Abbitt, J. T. (2011). Measuring technological pedagogical content knowledge in preservice teacher education: A review of current methods and instruments. *Journal of Research on Technology in Education*, 43(4), 281-300.

Abuzaghleh, O., Goldsmit, K., Elleithy, Y., y Lee, J. (2013.) Implementing an affordable high-performance computing for teaching-oriented computer science curriculum. *ACM Transactions on Computing Education (TOCE)*, vol. 13, no. 1, 1-14.

Aguar K, Arabnia H.R, Gutierrez J.B, Potter W.D, y Taha T.R. (2016). Making CS inclusive: An overview of efforts to expand and diversify CS education. In 2016 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 321-326). IEEE.

Aguar K, Safaei S, Arabnia H.R, Gutierrez J.B, Potter W.D, y Taha T.R (2017a.). Reviving Computer Science Education through Adaptive, Interest-Based Learning. In 2017 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1161-1166). IEEE.

Aguar, K., Sanchez, C.C., Beltran, D.B., Safaei, S., Asefi, M., Arnold, J., ... Gutierrez, J.B. (2017b). Considerations on Interdisciplinary Instruction and Design Influenced by Adaptive Learning. A Case Study Involving Biology, Computer Science, Mathematics and Statistics. arXiv preprint arXiv:1703.06010.

Akhgar, B., Saathoff, G. B., Arabnia, H. R., Hill, R., Staniforth, A. y Bayerl, P. S. (2015). Application of Big Data for National Security: A Practitioner's Guide to Emerging Technologies. Butterworth-Heinemann.

Albion, P., Jamieson-Proctor, R., y Finger, G. (2010). Auditing the TPACK Confidence of Australian Pre-Service Teachers: The TPACK Confidence Survey (TCS). In D. Gibson & B. Dodge (Eds.), Proceedings of SITE 2010--Society for Information Technology &

Teacher Education International Conference (pp. 3772-3779). San Diego, CA, USA: Association for the Advancement of Computing in Education (AACE). On-line December 28, 2018 from: <https://www.learntechlib.org/primary/p/33969/>.

Alonso-Pérez, S., Cuevas, E., Pérez, C., Querol, X., Baldasano, J. M., Draxler, R., y De Bustos, J. J. (2011). Trend changes of african airmass intrusions in the marine boundary layer over the subtropical Eastern North Atlantic region in winter. *Tellus Series B-Chemical and physical meteorology* 63(2): 255-265.

Amirian, S., Rasheed, K., Taha, T.R., y Arabnia, H. R. (2019). A Short Review on Image Caption Generation with Deep Learning. In Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV) (pp. 10-18). The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Anderson, A., Barham, N., y Northcote, M. (2013). Using the TPACK framework to unite disciplines in online learning. *Australasian Journal of Educational Technology*, 29 (4), 548-565. <http://dx.doi.org/10.14742/ajet.24>.

Anderson, L. M., Petticrew, M., Rehfuss, E. Armstrong, R., Ueffing, E., Baker, P., Francis, D., y Tugwell, P. (2011). Using logic models to capture Complexity in systematic reviews. *Research synthesis methods*, 2, 33-42.

Angeli, C., y Valanides, N. (2005). Preservice elementary teachers as information and communication technology designers: an instructional systems design model based on an expanded view of pedagogical content knowledge. *Journal of Computer Assisted Learning*, 21, pp.292-302. On-line: <http://onlinelibrary.wiley.com/doi/10.1111/jca.2005.21.issue-4/issuetoc>. Doi: 10.1111/j.13652729.2005.00135. x.

Arabnia, H. R. (1990). A Parallel Algorithm for the Arbitrary Rotation of Digitized Images using Process-and-Data-Decomposition Approach. *Journal of Parallel and Distributed Computing* 10(2): 188-193.

Arabnia, H. R. (1995). A Distributed Stereocorrelation Algorithm. *Proceedings of Computer Communications and Networks (ICCCN'95)*, IEEE: 479-482.

Arabnia, H. R., and Bhandarkar, S. M. (1996). Parallel Stereocorrelation on a Reconfigurable MultiRing Network. *The Journal of Supercomputing* (Springer Publishers) 10(3): 243-270.

Arabnia, H. R., y Oliver, M. A. (1987a). "A Transputer Network for the Arbitrary Rotation of Digitised Images". *The Computer Journal* 30(5): 425-433.

Arabnia, H. R., y Oliver, M. A. (1987b). "Arbitrary Rotation of Raster Images with SIMD Machine Architectures". *International Journal of Eurographics Association* (Computer Graphic Forum) 6(1): 3-12.

Arabnia, H. R., y Oliver, M. A. (1989). A Transputer Network for Fast Operations on Digitised Images. *International Journal of Eurographics Association* (Computer Graphics Forum) 8(1): 312.

Arabnia, H. R., y Smith, J. W. (1993). "A Reconfigurable Interconnection Network for Imaging Operations and Its Implementation Using a Multi-Stage Switching Box". *Proceedings of the 7th Annual International High Performance Computing Conference. The 1993 High Performance Computing: New Horizons Supercomputing Symposium*, Calgary, Alberta, Canada: 349-357.

Área, M. (2005). *La escuela y la sociedad de la información*. MEC Secretaría General Técnica. Nuevas tecnologías, globalización y migraciones. (pp. 13-54). Barcelona: Editorial Octaedro. On-line: <https://manarea.webs.ull.es/materiales/udtic/Escuela-SocInformacion.pdf>.

Arif, W. M., y Arabnia, H. R. (2003). Parallel Edge-Region-Based Segmentation Algorithm Targeted at Reconfigurable Multi-Ring Network. *The Journal of Supercomputing* 25 (1): 43-63.

- Armosky, B., Brown, S., Drumond T., Ferguson, J., Gerber, R., Hacker, T.J. y Traxler, K. (2007). HPC University. In TG08-TeraGrid Conference, Madison, WI.
- Asif, M., Cencerrado, A., Mula-Valls, O., Manubens, D., Doblas-Reyes, F., y Cortés, A. (2014). Impact of I/O and Data Management in Ensemble Large Scale Climate Forecasting Using ECEarth3. *Procedia Computer Science* 29: 2370-2379.
- Aspray, W., y Williams, B. O. (1994). Arming American scientists: NSF and the provision of scientific computing facilities for universities, 1950-1973. *Annals of the History of Computing*, IEEE 16(4): 60-74.
- Atkins, D. (2003). Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. National Science Foundation.
- Atkins, D., Borgman, C., Bindoff, N., Ellisman, M., Feldman, S., Foster, I., y Ynnerman A. (2010). Building a UK Foundation for the Transformative Enhancement of Research and Innovation. Report of the international panel for the 2009 review of the UK research councils e-science programme. Swindon, UK.
- Attig N., Gibbon, P., y Lippert, T. (2011). Trends in Supercomputing: The European path to exascale. *Computer Physics Communications* 182(9): 2041-2046.
- Aulkner, L. (2006). Internet 2 About us. URL: <http://www.internet2.edu/about>. 2014. Retrieved 12 Nov 2015.
- Avila, M., Folch, A., Houzeaux, G., Eguzkitza, B., Prieto, L., y Cabezón, D. (2013). A Parallel CFD Model for Wind Farms. *Procedia Computer Science* 18: 2157-2166.
- Bacon, J. (1998). *Concurrent System. Operating Systems, Database and Distributed Systems: an integrated approach*. Addison Wesley.
- Bader, D. A. (2004). Computational biology and high-performance computing. *Communications of the ACM* 47, 11, 34–41.
- Baker, T. L. (1994). *Doing social research*. 2nd ed. New York: McGrawHill.

- Balbinotti, M. A. A. (2004). Estou testando o que imagino estar? Reflexões acerca da validade dos testes psicológicos. *Técnicas projetivas: Produtividade em pesquisa*, 1, 22-26.
- Balbinotti, M. A. A., Benetti, M., y Soares Terra, R. (2007). Translation and validation of the Graham-Harvey survey for the Brazilian context. *International Journal of Managerial Finance*, 3(1), 26-48.
- Balladini, J., Grosclaude, E., Hanzich, M., Suppi, R., Rexachs del Rosario, D., y Luque Fadón, E. (2010). Incidence of Parallel Models and Scaling CPU Frequency in the Energy Consumption of the HPC Systems Programming In: XVI Congreso Argentino de Ciencias de la Computación (CACIC' 2010).
- Banerjee, P. (1994). *Parallel Algorithms for VLSI Computer-Aided Design*, Prentice-Hall.
- Baran, E., Chuang, H. H., and Thompson, A. (2011). TPACK: An emerging research and development tool for teacher educators. *TOJET*, 10(4).
- Barbara, M., y Donna, V. (2005). Learner-Centered Framework for E-learning. *Teachers college record*, 107(8), 1582-1600.
- Bazarganipour, F., Ziaei, S., Montazeri, A., Faghihzadeh, S., y Frozanfard, F. (2012). Psychometric properties of the Iranian version of modified polycystic ovary syndrome health-related quality-of-life questionnaire. *Human reproduction*, 27(9), 2729-2736.
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., and Esquivel, S. L. (2013). Practical Considerations for Using Exploratory Factor Analysis in Educational Research. *Practical Assessment, Research & Evaluation*. Volume 18, Number 6.
- Belfield, H. T., Bullock, A., Eynon, R., y Wall, D. (2001). Measuring effectiveness for best evidence medical education: a discussion. *Medical Teacher*, 23(2), 164-70.

Bergeron, K., Cummings, R., Decker, C.R., Freeman, M.J., Hoke, C.C., Seidel, J, Mc. y Daniel, D.M. (2008). An Undergraduate Computational Aerodynamics Curriculum. In DoD HPCMP Users Group Conference, 2008. DOD HPCMP UGC (pp. 18-24). IEEE.

Bermeo, H. P., De los Reyes, E., y Bonavia, T. (2008). Dimensions of the Scientific Collaboration and its Contribution to the Academic Research Groups' Scientific Quality. *Research Evaluation* 18(4): 301-311.

Bernabé, G., Cuenca, J., García, L.P., Giménez, D. y Rivas-Gómez, S. (2014). A High Performance Computing Course Guided by the LU Factorization. *Procedia Computer Science*, 29, 1446-1457.

Bernhard, A. (2009). A knowledge-based Society Needs Quality in Higher Education. *Problems of education in the 21st century* 12: 15-21.

Bethel, E. W., Van Rosendale, J., Southard, D., Gaither, K., Childs, H., Brugger, E., and Ahern, S. (2011). Visualization at Supercomputing Centers: The Tale of Little big Iron and the Three Skinny Guys. *IEEE Computer Graphics* 31(1): 90-95.

Bhandarkar, S. M., and Arabnia, H. R. (1995a). The Hough Transform on a Reconfigurable MultiRing Network. *Journal of Parallel and Distributed Computing* 24(1): 107-114.

Bhandarkar, S. M., and Arabnia, H. R. (1995). The REFINE Multiprocessor: Theoretical Properties and Algorithms. *Parallel Computing (journal)*, Elsevier, 21(11): 1783-1806.

Bhandarkar, S. M., Arabnia, H. R., and Smith, J. W. (1995). A Reconfigurable Architecture For Image Processing And Computer Vision. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)* (special issue on VLSI Algorithms and Architectures for Computer Vision, Image Processing, Pattern Recognition and AI) 9(2): 201-229.

Billingham, S., Whitehead, A. L., y Julious, S. A. (2013). An audit of sample sizes for pilot and feasibility trials being undertaken in the United Kingdom registered in the United Kingdom Clinical Research Network database, *BMC Medical Research Methodology* 13(104): 1–6. <http://dx.doi.org/10.1186/1471-2288-13-104>.

- Biswas, R., Dunbar, J., Hardman, J., Bailey, F. R., Wheeler, L., y Rogers, S. (2012). The Impact of High-End Computing on NASA Missions. *IT Professional* 14(2): 20-28.
- Bloom, B. (1994). *Reflections on the Development and Use of the Taxonomy*. In Anderson, L. Sosniak, L (Eds.) *Bloom's Taxonomy: A Forty-Year Retrospective*. Chicago: The National Society for the Study of Education, pp.1-8.
- Bollen, J., Fox, G., y Singhal, P. (2011). How and Where the TeraGrid Supercomputing Infrastructure Benefits Science. *Journal of Informetrics* 5(1): 114-121.
- Bomsdorf, B. (2005). Adaptation of learning spaces: Supporting ubiquitous learning in higher distance education. In *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum fr Informatik.
- Bornstedt, G.W. (1977). Reliability and Validity in Attitude Measurement. In: G.F. Summers (Ed.), *Attitude Measurement* (pp. 80-99). Kershaw Publishing Company: London.
- Bote-Lorenzo, M. L., Gómez-Sánchez, E., Vega-Gorgojo, G., Dimitriadis, Y. A., Asensio-Pérez, J. I., y Jorrín-Abellán, I. M. (2008). Gridcole: A tailorable grid service based system that supports scripted collaborative learning. *Computers & Education*, 51(1), 155-172.
- Brown, J. R., y Dinu, V. (2013). High performance Computing Methods for the Integration and Analysis of Biomedical Data using SAS. *Computer Methods and Programs in Biomedicine*, 112(3): 553-562.
- Bruce, B. C., Carragher, B. O., Damon, B. M., Dawson, M. J., Eurell, J. A., Gregory, C. D., ...y Potter, C. S. (1997). Chickscope: An interactive MRI classroom curriculum innovation for K-12. *Computers & Education*, 29(2-3), 73-87.
- Bryant, F. B., y Yarnold, P. R. (1995). Comparing five alternative factor-models of the Student Jenkins Activity Survey: separating the wheat from the chaff. *J Pers Assess.*,64: 145–158.

Bub, F. L., Mask, A. C., Wood, K. R., Krynen, D. G., Lunde, B. N., DeHaan, C. J., y Wallmark, J. A. (2014). The Navy's Application of Ocean Forecasting to Decision Support. *Oceanography*, 27(3): 126-137.

Cabero, A., Llorente, J., y Cejudo, M. C. (2013). La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información (TIC). Eduweb. *Revista de Tecnología de Información y Comunicación en Educación*, vol. 7, núm. 2, pp.11-22. On-line: <http://servicio.bc.uc.edu.ve/educacion/eduweb/v7n2/art01.pdf>.

Carey, P. (2013). Student as co-producer in a marketised higher education system: a case study of students' experience of participation in curriculum design. *Innovations in Education and Teaching International*, Vol. 50 , Iss. 3, 250-260.

Carretero, H., y Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International journal of clinical and healthpsychology*, 5(3):521-551.

Castle, T. L. (2008). Satisfied in the jail? Exploring the predictors of job satisfaction among jail officers. *Criminal Justice Review*, 33(1), 48-63.

Catlett, C., Allcock, W. E., Andrews, P., Aydt, R. A., Bair, R., Balac, N., y Marsteller, J. (2006). TeraGrid: Analysis of Organization, System Architecture, and Middleware Enabling New Types of Applications. High Performance Computing Workshop 16:225-249.

Cattell, R. B., and Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement*, 15(3), 139-164.

CERN - European Organization for Nuclear Research (2011) web, jul. 2015, URL: <http://public.web.cern.ch/public/>

Chai, C. S., Koh, J. H.-L., y Tsai, C. C. (2013a). A Review of Technological Pedagogical Content Knowledge. *Educational Technology & Society*, 16, 2, pp. 31-51.

- Chai, C.S., Ng, E.M.W., Li, W., Hong, H.Y., and Koh, J.H.L. (2013b). Validating and modeling technological pedagogical content knowledge framework among Asian pre-service teachers. *Australasian Journal of Educational Technology*, 29(1).
- Child, D. (2006). *The Essentials of Factor Analysis*. 3rd edn. New York: Continuum.
- Coates, H. (2005). The value of student engagement for higher education quality assurance. *Quality in Higher Education*, 11(1), 25–36.
- Comunidad Europea (2010): *Assessing Europe's University-Based Research*, 2010.
- Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1 (1), 104–126. URL <http://dx.doi.org/10.1007/BF03177550>
- Cosnard, M., y Trystran, D. (1995). *Parallel Algorithms and Architectures*. International Thomson Computer.
- Costello, A., y Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis. *Practical Assessment, Research & Evaluation*, Vol. 10 (7). On-line: <http://pareonline.net/genpare.asp?wh=0&abt=10>
- Coveney, P., McGuire, S., Parchment, O., Kenway, R., y Parsons, M. (2012). UK participation in high performance computing (HPC) at the european level - publications. URL: <https://www.gov.uk/government/publications/uk-participation-in-high-performance-computing-hpc-at-the-european-level>.
- Crocker, L., Llabre, M., y Miller, M. D. (1988). The generalizability of content validity ratings. *Journal of Educational Measurement*, 25(4), 287-299.
- Crouch, C.H. y Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American Journal of Physics*, vol. 69, no. 9, p. 970-977. URL: <http://link.aip.org/link/AJPIAS/v69/i9/p970/s1&Agg=doi>.

Davis, N. E., Robey, W., Ferenbaugh, C. R., Nicholaeff, D., y Trujillo, D. P. (2012). Paradigmatic shifts for exascale Supercomputing. *Journal of Supercomputing* 62(2): 1023-1044.

Dawes, R.M. (1975). *Fundamentos y técnicas de medición de actitudes*. México: Limusa.

De Benito, B., y Lizana, A. (2012). La formación de los docentes universitarios en TIC a partir de la transferencia del conocimiento de los docentes con experiencia en el uso de las TIC. In II Congreso Internacional EDO el (Vol. 23, p. 24).

De Filippo, D., Morillo, F., y Fernández, M. T. (2008). Indicadores de colaboración científica del CSIC con Latinoamérica en base de datos internacionales. [Indicators of Scientific Collaboration between CSIC and Latin America in International Databases]. *Revista Española de Documentación Científica*, 31(1): 66-84.

De Vicente, P. S. (1994). ¿Qué conocimiento necesitan los profesores? *Innovación Educativa*, 3, pp.11-31.

Denny, P., Hamer, J., Luxton-Reilly, A. y Purchase, H. (2008). PeerWise. In Proceedings of the 8th International Conference on Computing Education Research - Koli '08. New York, New York, USA: ACM Press, 2008, p. 109-112. URL: <http://portal.acm.org/citation.cfm?doid=1595356.1595378>.

Dongarra, J. J. y Van Der Steen, A. J. (2012). High-performance Computing Systems: Status and Outlook. *Acta Numérica* 21: 379-474.

Dongarra, J. J., Foster, I., Fox, G., y Gropp, W. (2002). *The Sourcebook of Parallel Computing*. Morgan Kaufmann.

Dongarra, J. J., Bunch, J. R., Moler, C. B., y Stewart, G. W. (1979). LINPACK users' guide. 8.

Edwards, A. L. (1957). *Techniques of attitude scale construction*. Nueva York: Appleton-Century-Croft.

Ellingson, S. R., Smith, J. C., y Baudry, J. (2014). Polypharmacology and Supercomputer-based Docking: Opportunities and Challenges. *Molecular Simulation*, 40(1): 10-11.

Elmagarmid, A. K., Samuel, A. y Ouzzani, M. (2008). Community-Cyberinfrastructure-Enabled Discovery in Science and Engineering. *IEEE Computing in Science and Engineering*, 10(5): 4653.

Emmott, S., y Rison, S. (2008). Towards 2020 science. *Science in Parliament*, 65(4), 31-33.

Escobar Pérez, J., y Cuervo Martínez, Á. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*, vol. 6, núm. 1, pp. 27-36. On-line: [http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3\\_Juicio\\_de\\_expertos\\_27-36.pdf](http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3_Juicio_de_expertos_27-36.pdf).

Fabricius, U., Freundl, C., Köstler, H. y Rüde, U. (2005) High performance computing education for students in Computational Engineering. In *Lecture Notes in Computer Science*, vol. 3515, no. II, pp. 27–35.

Farian, H., Anne, K.M. y Haas, M (2008). Teaching high-performance computing in the undergraduate college CS curriculum. *Journal of Computing Sciences in Colleges*, 23(3), 135-142.

Fawad, H., y Manarvi, I. A. (2014). Student feedback & systematic evaluation of teaching and its correlation to learning theories, Pedagogy & Teaching skills. In *Teaching, Assessment and Learning (TALE)*, International Conference on (pp. 398-404). IEEE.

Featherstone, N. A., Browning, M. K., Brun, A. S., y Toomre, J. (2009). Effects of Fossil Magnetic Fields on Convective Core Dynamos in A-type Stars. *The Astrophysical Journal* 705(1): 1000.

Fernández, Á., Fernández C., Miguel-Dávila, J.Á., Conde, M.Á, y Matellán, V. (2019). Supercomputers to improve the performance in higher education: A review of the literature. *Computers & Education*, 128, 353-364.

---

- Fernández Slezak, D, Turjanski, P.G., Montaldo, D. y Mocskos, E. E. (2010). Hands-On Experience in HPC with Secondary School Students. *IEEE Transactions on Education*, 53(1), 128-135.
- Fernbach, S. (1984). Supercomputers—Past, Present, Prospects. Future Generation. *Computer Systems*, 1(1): 23-30.
- Ferrando, P. J., y Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en Psicología. *Papeles del Psicólogo*, vol. 31, no. 1, pp. 18-33.
- Field, A. (2009). *Discovering Statistics using SPSS*. Sage: London.
- Fink, A. (2010). *Conducting Research Literature Reviews: From the Internet to Paper*. Sage Publications. 3rd edition.
- Foddy, W. (1996). *Constructing questions for interviews and questionnaires*. Cambridge: Cambridge University Press.
- Foster, I., Kesselman, C., y Tuecke, S. (2001). The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International Journal of High Performance Computing Applications*, 15(3): 200–222.
- Frías-Navarro, D., y Pascual Soler, M. (2012). Prácticas del análisis factorial exploratorio (AFE) en la investigación sobre conducta del consumidor y marketing. *Suma Psicológica*, vol. 19, no. 1, pp. 47-58.
- Gable, R. K., y Wolf, M. B. (1986). *Instrument Development in the Affective Domain*. Boston: Kluwer-Nijhoff.
- Gaudiani, A. (2012). Análisis del rendimiento de algoritmos paralelos de propósito general en GPGPU [Performance Analysis of Parallel Algorithms for General Purposes in GPGPU] (Doctoral dissertation, Facultad de Informática). URL: [http://sedici.unlp.edu.ar/bitstream/handle/10915/22691/Documento\\_completo\\_\\_\\_\\_.pdf?sequence=1](http://sedici.unlp.edu.ar/bitstream/handle/10915/22691/Documento_completo____.pdf?sequence=1) Retrieved 2015 10 08.

Geller, T. (2011). Supercomputing's exaflop target. *Communications of the ACM*, 54(8): 16-18.

Gengler, M., Ubeda, S., y Desprez, F. (1996). "Initiation au parallélisme: concepts, architectures et algorithmes" [Introduction to Parallelism: Concepts, Architectures and Algorithms]. Masson ISBN 2-225-85014-3.

George, E. I. (2000). The variable selection problem. *Journal of the American Statistical Association*, 95(452), 1304-1308.

Gliem, J., y Gliem, R. (2003). Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. Conference in Adult, Continuing, and Community Education. Midwest Research to Practice.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.

González, Á. F., Rosillo, R., Dávila, J. Á. M., and Olivera, V. M. (2015). Historical review and future challenges in Supercomputing and Networks of Scientific Communication. *Journal of Supercomputing*. 71(12), 4476-4503. ISSN 0920-8542.

Graham, C. R., Burgoyne, N., Cantrell, P., Smith, L., St. Clair, L., y Harris, R. (2009a). TPACK Development in Science Teaching: Measuring the TPACK Confidence of Inservice Science Teachers, *TechTrends*, Special Issue on TPACK, 53(5), 70-79.

Graham, C., Cox, S., ay Velasquez, A. (2009b). Teaching and Measuring TPACK Development in Two Preservice Teacher Preparation Programs. In I. Gibson, R. Weber, K. McFerrin, R. Carlsen & D. Willis (Eds.), *Proceedings of SITE 2009--Society for Information Technology & Teacher Education International Conference* (pp. 4081-4086). Charleston, SC, USA: Association for the Advancement of Computing in Education (AACE). On-line:<https://www.learntechlib.org/primary/p/31297/>. Retrieved: december 2018.

- Grossman, P. L. (1990). A study in contrast: Sources of pedagogical content knowledge for secondary English. *Journal of Teacher Education*, 40(5), 24-31.
- Guellec, D. y Van Pottelsberghe de la Potterie, B. (2001). R&D and Productivity Growth: Panel Data Analysis of 16 OECD Countries. STI Working Papers 2001/3. OECD.
- Guevara, M., Martínez, F., Arevalo, G., Gassó, S., y Baldasano, J. (2013). An Improved System for Modelling Spanish Emissions: HERMESv2.0. *Atmospheric Environment*, 81: 209-22.
- Gülbahar, Y. (2005). Technology Planning: A Roadmap to Successful Technology Integration in Schools. *Computers & Education*, 49(4), 943-956.
- Gutiérrez, R. C., Zagalaz, J., y López, J. M. S. (2015). Creando contenidos curriculares digitales de Ciencias Sociales para Educación Primaria. Una experiencia TPACK para futuros docentes. *Educatio Siglo XXI*, 33(3), 147-168.
- Hacker, T. (2010). Hands-on high performance computing developing a cluster computing course for real world supercomputing. In Proceedings of the National ASEE Conference.
- Hair, J., Anderson, R. E., Tatham, R. L., y Black, W. (1995). *Multivariate data: Analysis with readings*. New Jersey: Prentice-Hall.
- Hammick, M. (2005). A BEME review: a little illumination. *Medical Teacher*, 27(1), 1-3.
- Hammick, M., Dornan, T., y Steinert, Y. (2010). Conducting a best evidence systematic review. Part 1: From idea to data coding. BEME Guide No. 13. *Medical Teacher*, 32(1), 3-15.
- Harden, M., Grant, J., Buckley, G., and Hart, I.R. (1999). BEME Guide No. 1: Best evidence medical education. *Medical Teacher*, 21(6), 553-62.
- Hassan, A. H., Fluke, C. J. and Barnes, D. G. (2011). Unleashing the Power of Distributed CPU/GPU architectures: Massive Astronomical Data Analysis and Visualization Case Study. arXiv preprint arXiv:1111.6661.

Hauben, M. (2010). History of ARPANET. URL: <http://pages.infinet.net/jbcoco/Arpa-ArpanetInternet.pdf> Retrieved 2015.

Hazzi, O., y Maldaon, I. (2015). A pilot study: Vital methodological issues. *Business: Theory and Practice*, 16, 53.

Hemanth, J.D., Kose, U., Deperlioglu, O. y De Albuquerque, V.H.C. (2012). An augmented reality-supported mobile application for diagnosis of heart diseases. *The Journal of Supercomputing*, 76 (2), 1242-1267.

Hermes, H. (1969). Enumerability, Decidability, Computability. Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen. Band 127. Springer-Verlag. Berlin, Heidelberg, New York, 1969.

Hernández-Nieto, R. A. (2002). Contribuciones al análisis estadístico. Mérida, Venezuela: Universidad de Los Andes/IESINFO.

High Performance Computing Act-HPCA (1991). Act of Congress promulgated in the 102nd United States Congress as (Pub.L. 102–194) on 1991-12-09.

Holmes, V. y Kureshi, I. (2015) Developing High Performance Computing Resources for Teaching Cluster and Grid Computing Courses. *Procedia Computer Science*, 51, 1714-1723.

Hooper, B., King, R., Wood, W., Bilics, A., y Gupta, J. (2013). An international systematic mapping review of educational approaches and teaching methods in occupational therapy. *The British Journal of Occupational Therapy*, 76(1), 9-22.

Huberty, C. J. (1989). Problems with stepwise methods: Better alternatives. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 1, pp. 43-70). Greenwich, CT: JAI Press.

Huberty, C. J. (1994). *Applied discriminant analysis*. New York: Wiley and Sons.

Hunter, J. E., Schmidt, F. L., y Jackson, G. B. (1982). *Meta-analysis: Cumulating Research Findings across Studies*, Sage, Beverly Hills.

Hwang, K. (1993). *Advanced Computer Architecture: Parallelism, Scalability, Programmability*. McGraw-Hill 1993.

IDC (2004). White Paper: Council on Competiveness Study of U.S. Industrial HPC Users. URL: [http://www.compete.org/storage/images/uploads/File/PDF%20Files/HPC\\_Users\\_Survey%202004.pdf](http://www.compete.org/storage/images/uploads/File/PDF%20Files/HPC_Users_Survey%202004.pdf).

Infante, E. F. (1986). *An Agenda for Improved Evaluation of Supercomputer Performance: A Report*. National Academies Press.

Joiner, D. A., Gray, P., Murphy, T., y Peck, C. (2006). Teaching parallel computing to science faculty: Best practices and common pitfalls. In *Proceedings of the 11th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP '06)*. ACM, New York, 239–246.

Jonassen, D. H. (1999). *Computers as mindtools for schools: Engaging critical thinking*. NJ: Prentice Hall, Inc.

Joshua, R. y Scuse, D.H, (2005), Modeling beliefs and solution strategies in a distributed learning system. *The Journal of Supercomputing*, 34(1), 27-39.

Kantardjiev, A. A. (2012). Quantum.Ligand.Dock: protein-ligand docking with quantum entanglement refinement on a GPU system. *Nucleid Acids Research*, 40(W1): W415-W422.

Kennedy, K., y Joy, W. (1998). *Interim Report to the President, President's Information Technology Advisory Committee (PITAC)*. National Coordination Office for Computing, Information and Communication, 4201 Wilson Blvd, Suite 690, Arlington, VA 22230, August 10, 1998.

- Kim, K. y Moon, N. (2019). Activity index model for self-regulated learning with learning analysis in a TEL environment. *The Journal of Supercomputing*, 75(4), 1971-1989.
- Kim, J. O., and Mueller, C. W. (1978). *Factor Analysis, Statistical Methods and Practical Issues*. Sage, Beverly Hills.
- Kinter, J. L., Cash, B., Achuthavarier, D., Adams, J., Alshuler, E., Dirmeyer, P., y Wong, K. et al. (2013). Revolutionizing Climate Modeling with Project Athena. A Multi-Institutional, International Collaboration. *Bulletin of the American Meteorological Society* 94(2): 231-245.
- Kirkpatrick, D. I. (1967). *Evaluation of training*. In: R. Craig, I Mittel, eds. Training and development handbook. New York: McGraw Hill.
- Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. Joint Technical Report, Software Engineering Group, Keele University, and Empirical Software Eng., Nat'l ICT Australia.
- Kline, P. (1999). *The Handbook of Psychological Testing*. London: Routledge.
- Koehler, M. J., y Mishra, P. (2005). What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *Journal of Educational Computing Research*, 32(2), 131-152.
- Kramer, D. (2011). Supercomputing has a future in clean energy. *Physics Today*, 64(7): 27-29.
- Kupczyk, M., y Meyer, N. (2010). PRACE World-Class Computational Facilities Ready for Polish Scientific Community. *Computational Methods in Science and Technology*, 2010 57-62.
- Kwon, H. I., Kim, S., Lee, H., Ryu, M., Kim, T., and Choi, S. (2013). Development of an Engineering Education Framework for Aerodynamic Shape Optimization. *International Journal of Aeronautical and Space Sciences*, 14(4): 297-309.

- Lavington, S. H. (1978). The Manchester Mark I and Atlas: a Historical Perspective. *Communications of the ACM*, 21(1): 4-12.
- Lawson, S. J., Woodgate, M., Steijl, R., y Barakos, G. N. (2012). High Performance Computing for Challenging Problems in Computational Fluid Dynamics. *Progress in Aerospace Sciences*, 52: 19-29.
- Lederman, D., y Maloney, W. (2003). R&D and Development. World Bank Policy Research Working Paper 3024.
- Lee, C.M. y Chang, H. (2014). A study on security strategy in ICT convergence environment. *The Journal of Supercomputing*, 70(1), 211-223.
- Lee, E., Huang, H., Dennis, J. M., Chen, P. y Wang, L. (2013). An Optimized Parallel LSQR Algorithm for Seismic Tomography. *Computer & Geosciences*, 61: 184-197.
- Leedy, P. D., y Ormrod, J. E. (2001). *Practical Research: Planning and Design*, seventh ed., Merrill Prentice Hall, Upper Saddle River, NJ.
- Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., y Wolff, S. (2009). A Brief History of the Internet. *ACM SIGCOMM Computer Communication Review*, 39(5), 22-31.
- Lepora, N. F., Overton, P. G., y Gurney, K. (2012). Efficient Fitting of Conductance-based Model Neurons from Somatic Current Clamp. *Journal of Computational Neuroscience*, 32(1): 1-24.
- Levy, Y., y Ellist, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research, *Inf. Sci.*, 9, pp. 181–212.
- Lim, C. P. (2005). Online learning in higher education: necessary and sufficient conditions. *International Journal of Instructional Media*, 32(4), 323.

- Lingerfelt, E. J., Messer, O. E., Desai, S. S., Holt, C. A., y Lentz, E. J. (2014). Near Real-time Data Analysis of Core-collapse Supernova Simulations with Bellerophon. *Procedia Computer Science*, 29: 1504-1514.
- Lu, F., Song, J., Cao, X. y Zhu, X. (2012). CPU/GPU computing for long-wave radiation physics on large GPU clusters. *Comput. Geosci.* 41, 47–55.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing research*.
- MacCallum, R. C., Widaman, K. F., Zhang, S., y Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84–99.
- Markram, H., Meier, K., Lippert, T., Grillner, S., Frackowiak, R., Dehaene, S., y Saria, A. (2011). Introducing the Human Brain Project. *Procedia Computer Science*, 7: 39-42.
- Martín, A.G. (2007). Integración curricular de las TIC y educación para los medios en la sociedad del conocimiento. *Revista Iberoamericana de educación*, 45, 141-156.
- Martin, H. S. C., Jha, S., y Coveney, P. V. (2014). Comparative Analysis of Nucleotide Translocation through Protein Nanopores Using Steered Molecular Dynamics and an Adaptive Biasing Force. *Journal of Computational Chemistry*, 35(9): 692-702.
- Matsuzaki Y., Uchikoga, N., Ohue, M., Shimoda, T., Sato, T., Ishida, T., y Akiyama, Y. (2013). MEGADOCK 3.0: A High-performance Protein-protein Interaction Prediction Software using Hybrid Parallel Computing for Petascale Supercomputing Environments. *Source code for biology and medicine*, 8: 18.
- Mc. Cartney, S. (1999). ENIAC: The Triumphs and Tragedies of the World's First Computer. Walker & Company. ISBN:0802713483.
- Medina Molina, M., Castro Sánchez, J. J., Viera-Santana, J. G., y García-Quevedo, E. (2014). Tecnologías para el acceso a internet que utilizan los estudiantes de la ULPGC.

- Menhorn, F. y Reumann, M. (2013). Genome Assembly Framework on Massively Parallel, Distributed Memory Supercomputers. *Biomedical Engineering/Biomedizinische Technik* 2013.
- Merriam, S.B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass Publishers, 350 Sansome St, San Francisco, CA 94104.
- Metcalf, T. S., Mathur, S., Dogan, G., y Woitaszek, M. (2012). First Results from the Asteroseismic Modeling Portal. *Progress in Solar/Stellar Physics with Helio-and Asteroseismology*, 462: 213.
- Meuer, H. W., y Gietl, H. (2013). Supercomputers—Prestige Objects or Crucial Tools for Science and Industry? *PIK-Praxis der Informationsverarbeitung und Kommunikation*, 36(2): 117-128.
- Mikropoulos, T. A., y Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, 56(3), 769-780.
- Milone, D., Azar, A., y Rufiner, L. (2002). Supercomputadoras basadas en "clusters" de PCs. Trabajo de desarrollo tecnológico realizado en el Laboratorio de Cibernética de la Facultad de Ingeniería (UNER). [Supercomputers based on PC "clusters." Technological Development Project Carried out in the Cybernetic Laboratory of the Engineering Faculty]. *Revista Ciencia, Docencia y Tecnología*, Vol. XIII, No. 25, pp. 173-208, nov. 200268.
- Mohammadi, F.G., Amini, M.H. y Arabnia, H.R. (2020). Evolutionary Computation, Optimization and Learning Algorithms for Data Science. In *Optimization, Learning, and Control for Interdependent Complex Networks* (pp. 37-65). Springer, Cham.
- Moraleda, A. (2007). Supercomputing: a Qualitative Leap for Competitiveness. *Economistas*, 26(116): 294-297.
- Morales, P. (1988). *Medición de actitudes en psicología y educación. Construcción de escalas y problemas metodológicos*. San Sebastián: Ttarttalo.

Munetomo, M. (2011). Realizing Robust and Scalable Evolutionary Algorithms toward Exascale era. *Evolutionary Computation (CEC), 2011 IEEE Congress on* (pp. 312-317). IEEE.

Munshi, R., Coalson, R. D., Ermentrout, G.B., Madura, J.D., Meirovitch, H., Stiles, J. R. y Bahar, I. (2006). An introduction to simulation and visualization of biological systems at multiple scales: a summer training program for interdisciplinary research. *Biotechnology progress*, 22(1), 179-185.

Muñoz-Repiso, A. G. V., y del Pozo, M. M. (2016). Análisis de las competencias digitales de los graduados en titulaciones de maestro/Analysis of the digital competences of graduates of university degrees to be a teacher. *Revista Latinoamericana de Tecnología Educativa-RELATEC*, 15(2), 155-168.

Murtaugh, P. A. (2009). Performance of several variable-selection methods applied to real ecological data. *Ecology letters*, 12(10), 1061-1068.

Mustafa, K. O. Ç. (2005). Implications of learning theories for effective technology integration and pre-service teacher training: A critical literature review. *Journal of Turkish Science Education*, 2(1), 2.

National Academy of Sciences (2005). Getting up to Speed the Future of the Future of Supercomputing. Committee on the Future of Supercomputing Computer Science and Telecommunications Board Division on Engineering and Physical Sciences.

Neic, A., Liebmann, M., Hoetzl, E., Mitchell, L., Vigmond, E. J., Haase, G., y Plank, G. (2012). Accelerating Cardiac Biodomain Simulations Using Graphics Processing Units. *IEEE. Transactions on biomedical engineering*, 59(8): 2281-2290.

Nesmachnow, S. (2014). Planificación de tareas en sistemas cluster, grid y cloud utilizando algoritmos evolutivos. [Scheduling in Cluster Systems, Grids and Clouds using Evolutionary Algorithms]. *Revista de la Sociedad Mexicana de Inteligencia Artificial*, 18 / 36 1: Jan - April 2014.

- Nicolosi, F., Della Vecchia, P., y Ciliberti, D. (2013). An Investigation on Vertical Tail Plane Contribution to Aircraft Sideforce. *Aerospace Science and Technology*, 28(1): 401-416.
- Nunnally, J. C. y Bernstein, I. H. (1994). *Psychometric Theory*, third edition. New York: McGraw-Hill.
- Nunnally, J. C. (1978). *Psychometric theory*. 1. New York: McGraw-Hill.
- OECD (2015). Students, Computers and Learning, 2015. On-line: <http://dx.doi.org/10.1787/9789264239555-en>.
- O'Neill, J. E. (1995). The Role of ARPA in the Development of the ARPANET, 1961-1972. *Annals of the History of Computing, IEEE*, 17(4): 76-81.
- Osborne, J. W., and Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Pract Assess Res Eval* [Internet]. On-line: <http://PAREonline.net/getvn.asp?v=9&n=11>.
- Patterson, C. A., Snir, M., and Graham, S. L. (2005). *Getting Up to Speed: The Future of Supercomputing*. National Academies Press.
- Perine, K. (2000). The Early Adopter - Al Gore and the Internet - Government Activity. The Industry Standard.
- Pierson, M. E. (2001). Technological integration practice as a function of pedagogical expertise. *Journal of Research on Computing in Education*, 33, pp. 413-430.
- Plichta, S. B., Kelvin, E. A., y Munro, B. H. (2013). *Munro's statistical methods for health care research*. Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Porta, M. (2008). *A dictionary of epidemiology*. 5th ed. Oxford: Oxford University Press. 320 p.

Puckelwartz, M. J., Pesce, L., Nelakuditi, V., Dellefave-Castillo, L., Golbus, J. R., Day, S. M. y McNally, E. M. (2014). Supercomputing for the Parallelization of Whole Genome Analysis. *Bioinformatics*, 30(11): 1508-1513.

Rattray, J.C., y Jones, M.C. (2007), Essential elements of questionnaire design and development. *Journal of Clinical Nursing*, 16, 234-243.

RedCLARA (2011). Compendio RedCLARA de Redes Nacionales de Investigación y Educación Latinoamericanas. [CLARA Compendium of Latin American National Research Networks and Education]. Bogotá: RedCLARA.

Rennie, K. M. (1997). Exploratory and Confirmatory Rotation Strategies in Exploratory Factor Analysis. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, January, 1997.

Reumann, M., Makalic, E., Goudey, B. W., Inouye, M., Bickerstaffe, A., Bui, M., y Hopper, J. L. (2012). Supercomputing Enabling Exhaustive Xstatistical Analysis of Genome wide Association Study Data: Preliminary results. Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE (pp. 1258-1261). IEEE.

Romer, P. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98(5): 71102.

Rosenberg, L. C. (1991). Update on National Science Foundation funding of the “Collaboratory”. *Communications of the ACM*, 34(12): 83.

Ryabinin, K. y Chuprina, S. (2015). Using Scientific Visualization Tools to Bridge the Talent Gap. *Procedia Computer Science*, 51, 1734-1741.

Saini, S., Rappleye, J., Chang, J., Barker, D., Mehrotra, P., y Biswas, R. (2012). I/O Performance Characterization of Lustre and NASA Applications on Pleiades. High Performance Computing (HiPC), 2012 19th International Conference on (pp. 1-10). IEEE.

- Sanz, M. A. (1998). Fundamentos históricos de la Internet en Europa y en España. [Historical Foundations of the Internet in Europe and Spain]. Boletín Rediris 45
- Saunders, V. R. y Guest, M. F. (1982). Applications of the CRAY-1 for quantum chemistry calculations. *Computer Physics Communications*, 26(3): 389-395.
- Sawyer, M., y Parsons, M. (2011). A Strategy for Research and Innovation through High Performance Computing. The University of Edinburgh 2011.
- Schaller, R. (1997). Moore's Law: Past, Present and Future. *Spectrum, IEEE*, 34(6): 52-59.
- Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., y Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK) the development and validation of an assessment instrument for preservice teachers. *Journal of research on Technology in Education*, 42(2), 123-149.
- Semiz, K., e Ince, M.L. (2012). Pre-service physical education teachers' technological pedagogical content knowledge, technology integration self-efficacy and instructional technology outcome expectations. *Australasian Journal of Educational Technology*, 28(7), 1248–1265.
- Shen, B., Nelson, B., Cheung, S. y Tao, W. K. (2013). Improving NASA's Multiscale Modeling Framework for Tropical Cyclone Climate Study. *Computing in Science & Engineering*, 15(5): 56-67.
- Shiflet, A.B. y Shiflet, G.W. (2010). Simulating the formation of biofilms in an undergraduate modeling course. *Procedia Computer Science*, 1(1), 895-901.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15 (2), 4–14.
- Simmons, M. L., y Wasserman, H. J. (1990). Performance comparison of the CRAY-2 and CRAY X-MP/416 supercomputers. *Journal of Supercomputing*, 4(2): 153-167.

Singh, S. (2000). *The Code Book: the Secret History of Codes and Code-breaking: 77-85*. London: Fourth Estate.

Skupin, A., Biberstine, J., y Boerner, K. (2013). Visualizing the Topical Structure of the Medical Sciences: A Self-Organizing Map Approach. *Plos One*, 8(3).

Smeets, E. (2005). Does ICT contribute to powerful learning environments in primary education?. *Computers & Education*, 44 (3), 343–355.

Smith, N. J., y Bhattacharya, K. (2015). Practical Wisdom of Tool and Task.

Smith, J., y Wolffe, G. (2007). Introducing AP computer science students to high-performance computing. *Journal of Computing Sciences in Colleges*, 23(1), 70-76.

Soete, L., O'Doherty, D., Arnold, E., Bounfour, A., Fagerberg, J., Farinello, U., y Schiestock, G. (2002). Benchmarking National Research Policies: The Impact of RTD on Competitiveness and Employment (IRCE). A Strata-ETAN Expert Working Group, European Commission DG Research, Brussels.

Specht, M., Bedek, M., Duval, E., y Held, P. (2013). WESPOT: Inquiry based learning meets learning analytics. Retrieved from: <http://bg-openaire.eu/handle/10867/130>.

Stainsby, H., Muresano, R., Fialho, L., González, J.C., Rexachs, D. y Luque, E. (2009). Teaching model for computational science and engineering programme. In *Computational Science-ICCS 2009* (pp. 34-43). Springer Berlin Heidelberg.

Stake, R.E. (1994). *Case studies*. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 236–247). Thousand Oaks, CA: Sage.

Stamatakis, A., Aberer, A. J., Goll, C., Smith, S. A., Berger, S. A., y Izquierdo-Carrasco, F. (2012). RAxML-Light: a Tool for Computing Terabyte Phylogenies. *Bioinformatics*, 28(15): 2064-2066.

Stein, R. F., Lagerfjård, A., Nordlund, Å., y Georgobiani, D. (2012). Helioseismic Data from Emerging Flux Simulations. *Progress in Solar/Stellar Physics with Helio-and Asteroseismology*, 462: 345.

Stewart, A. L., y Ware, J. F. (1992). *Measuring functioning and well-being: the medical outcome study approach*. Durham: Duke University Press, pp. 76-81.

Stojanovic, N. y Milanovic, E. (2015). Teaching Introductory Parallel Computing Course with Hands-On Experience. *International Journal of Engineering Education*, 31(5), 1343-1351.

Stone, H. (1993). *High Performance Computer Architectures*. Addison Wesley.

Stoyanov, D., Grigorov, I., Deleva, A., Kolev, N., Peshev, Z., Kolarov, G., e Ivanov, D. (2013). Remote Monitoring of Aerosol Layers over Sofia during Sahara Dust Transport Episode (April, 2012). In Seventeenth International School on Quantum Electronics: Laser Physics and Applications (pp. 87700Y-87700Y). International Society for Optics and Photonics.

Suárez, T.L. (2007) Access Grid technology in classroom and research environments. *The Journal of Supercomputing*, 41(2), 133-145.

Sumiyoshi, K. (2011). A numerical Challenge on the Core-collapse Supernovae: Physics of Neutrino and Matter at Extreme Conditions. *Journal of Physics: Conference Series* 302(1).

Sung, N. S., Gordon, J. I., Rose, G. D., Getzoff, E. D., Kron, S. J., Mumford, D., Onuchic, J. N., Scherer, N. F., ...y Kopell, N. J. (2003). Educating future scientists.

Thompson, B. (1989). Why won't stepwise methods die? *Measurement and Evaluation in Counseling and Development*, 21 (4), 146-148.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington DC: American Psychological Association.

- Thornton, J. E. (1970). Design of a Computer—The control data 6600.
- Thurstone, L. L. (1947). Multiple-factor analysis; a development and expansion of The Vectors of Mind.
- Tochel, C., Haig, A., Hesketh, A., Cadzow, A., Beggs, K., Colthart, I., y Peacock, H. (2009). The effectiveness of portfolios for post-graduate assessment and education: BEME Guide No. 12. *Medical Teacher*, 31(4), 299-318.
- Toomre, J., Augustson, K. C., Brown, B. P., Browning, M. K., Brun, A. S., Featherstone, N. A., and Miesch, M. S. (2012). New Era in 3-D Modeling of Convection and Magnetic Dynamos in Stellar Envelopes and Cores. *Progress in Solar/Stellar Physics with Helio-and Asteroseismology*, 462(331).
- Tran, Q.N. y Arabnia, H. (2015). *Emerging trends in computational Biology, Bioinformatics and System Biology*. Moran Kauffmann publishers, Elsevier.
- Turpin, T., Lian, Y., Tong, J., y Fang, X. (1995). Technology and Innovation Networks in the People's Republic of China. *Journal of Industry Studies*, 2(2): 63-74.
- Um, J., Choi, H., Song, S. K., Choi, S. P., Mook Yoon, H., Jung, H., and Kim, T. H. (2013). Development of a Virtualized Supercomputing Environment for Genomic Analysis. *Journal of Supercomputing*, 65(1): 71-85.
- Utreras, F. (2014). Visión y Proyectos: el presente y el futuro. [Vision and Projects: Present and Future]. XI Encuentro Temático Nacional Renata – RUP. RedClara.
- Villarubia, C. (2012). “La Comisión Europea dobla el presupuesto para HPC” [“The European Commission Doubles the Budget for HPC”]. URL: <http://www.bsc.es/sites/default/files/public/about/news/hpc-22022012-datacenterdynamics.pdf>, Retrieved: 27/08/2014.
- Volman, M. (2005). A variety of roles for a new type of teacher educational technology and the teaching profession. *Teaching and Teacher Education*, 21(1), 15-31.

Von Neumann, J. (1945). First Draft of a Report on the EDVA. Between the United States Army Ordinance Department and the University of Pennsylvania Moore School of Electrical Engineering University of Pennsylvania. June 30, 1945. Contract No. W-670-ORD-4926.

Wagner, J. M., y Shimshak, D. G. (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *European Journal of Operational Research*, 180(1), 57-67.

Wang, R., y Harris, C. (2013). Scaling Radio Astronomy Signal Correlation on Heterogeneous Supercomputers using Various Data Distribution Methodologies. *Experimental Astronomy*, 36(3): 433-449.

Wang, A. Y. (2013). The Impact of Digital Storytelling on the Development of TPACK among Student Teachers. In Society for Information Technology & Teacher Education International Conference (Vol. 2013, No. 1, pp. 1794-1794).

Wen, J., Li, S., Lin, Z., Hu, Y., y Huang, C. (2012). Systematic literature review of machine learning based software development effort estimation models. *Information and Software Technology*. 54 (1), 41 – 59.

Westerlund, S., y Harris, C. (2014). A Framework for HI Spectral Source Finding Using Distributed-Memory Supercomputing. *Publications of the Astronomical Society of Australia*, 31, e023 doi:10.1017/pasa.2014.18.

Whitaker, J. S. (1997). Use of Stepwise Methodology in Discriminant Analysis.

Wilkes, M. V., y Renwick, W. (1950). The EDSAC (Electronic Delay Storage Automatic Calculator). *Mathematics of Computation*, 4(30): 61-65.

Winkel, M., Speck, R., Huebner, H., Arnols, L., Krause, R., y Gibbon, P. (2012). A Massively Parallel, Multi-disciplinary Barnes-Hut Tree Code for Extreme-scale N-body Simulations. *Computer Physics Communications*, 183(4): 880-889.

- Woolgar, L. (2007). New Institutional Policies for University-Industry Links. *Japan Research Policy*, 36: 1261-1274.
- Wu, W. (2007). Cultivating Research Universities and Industrial Linkages in China: The Case of Shanghai. *World Development*, 35(6):1075-1093.
- Wu, Y., y Cai, X. C. (2014). A fully Implicit Domain Decomposition based ALE Framework for Three-dimensional Fluid-structure Interaction with Application in Blood Flow Computation. *Journal of Computational Physics*, 258: 524-537.
- Wyld, M. (1996) Technology refusa: Rationalising the failure of student and beginning teachers to use computers. *British Journal of Educational Technology*, 27, 134–143.
- Wynd, C. A., Schmidt, B., y Schaefer, M. A. (2003): Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research*, 25, pp. 508–518.
- Yamamoto, K. (2004). Corporatization of National Universities in Japan: Revolution for Governance or Rhetoric for Downsizing?. *Financial Accountability & Management*, 20(2): 153-181.
- Yang, M.Q., Yu, S., Yang, W., Milanova, M., Zhao, W., Yang, J.Y. y Arabnia, H.R. (2017). Developing Secure Privacy Preserving and Causal Genetic Alteration Research in Building an Innovative Systematic Pedagogy for Integrated Research and Education: The INSPIRE Model. In 2017 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1149-1154). IEEE.
- Yang, W., Yoshigoe, K., Qin, X., Liu, J. S., Yang, J. Y., Niemierko, A., Deng, Y., Liu, Y., Dunker, A. K., Chen, Z., Wang, L., Xu, D., Arabnia, H. R., Tong, W., y Yang, M. Q. (2014). Identification of Genes and Pathways Involved in Kidney Renal Clear Cell Carcinoma. *BMC bioinformatics*, 15(Suppl 17), S2.
- Yang, C. H., Tzuo, P. W., y Komara, S. (2011). WebQuest and collaborative learning in teacher preparation. Singapore study. *Educational Media International*, 48(3), 209-220.

- Yasar, O., Maliekal, J., Little, L.J. y Jones, D. (2006). A computational technology approach to education. *Computing in Science and Engineering*, 8(3), 76-81.
- Yeun, E. J., y Shin-Park, K. K. (2006). Verification of the profile of mood states-brief: Cross-cultural analysis. *Journal of Clinical Psychology*, 62(9), 1173-1180.
- Yin, R.K. (2009). *Case study research: Design and methods* (4th ed.). London: Sage.
- Yin, S., Luo, H., and Ding, S. (2013). Real-time Implementation of Fault-tolerant Control Systems with Performance Optimization. *Industrial Electronics, IEEE Transactions*, 61(5): 2402 – 2411.
- Yoshida, H. (2013). Cloud-super-computing Virtual Colonoscopy with Motion-based Navigation for Colon Cancer Screening. Consumer Electronics?? Berlin (ICCE-Berlin), 2013. ICCE Berlin 2013. IEEE Third International Conference on (pp. 165-167). IEEE.
- Young, J. R., Young, J. L., y Shaker, Z. (2012). Technological Pedagogical Content Knowledge (TPACK). Literature Using Confidence Intervals. *TechTrends*.
- Zhou, P., y Leydesdorff, L. (2006). The Emergence of China as a Leading Nation in Science. *Research Policy*, 35(1): 83-104.