AJUR

American Journal of
Undergraduate Research

Volume 18 | Issue 2 | September 2021

www.ajuronline.org

Print Edition ISSN 1536-4585
Online Edition ISSN 2375-8732

AJUR
American Journal of
Undergraduate Research

Volume 18 | Issue 2 | September 2021 *www.ajuronline.org*

American Journal of Undergraduate Research (AJUR) *is a national, independent, peer-reviewed, open-source, quarterly, multidisciplinary student research journal. Each manuscript of AJUR receives a DOI number. AJUR is archived by the US Library of Congress. AJUR was established in 2002, incorporated as a cheritable not-for-profit organization in 2018. AJUR is indexed internationally by EBSCO and Crossref with ISSNs of 1536-4585 (print) and 2375-8732 (web).*

On Packing Thirteen Points in an Equilateral Triangle

*Natalie Tedeschi**

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA

https://doi.org/10.33697/ajur.2021.042

Student: ntedesch@andrew.cmu.edu
Mentor: jmackey@andrew.cmu.edu

ABSTRACT
The conversation of how to maximize the minimum distance between points - or, equivalently, pack congruent circles - in an equilateral triangle began by Oler in the 1960s. In a 1993 paper, Melissen proved the optimal placements of 4 through 12 points in an equilateral triangle using only partitions and direct applications of Dirichlet's pigeon-hole principle. In the same paper, he proposed his conjectured optimal arrangements for 13, 14, 17, and 19 points in an equilateral triangle. In 1997, Payan proved Melissen's conjecture for the arrangement of fourteen points; and, in September 2020, Joos proved Melissen's conjecture for the optimal arrangement of thirteen points. These proofs completed the optimal arrangements of up to and including fifteen points in an equilateral triangle. Unlike Melissen's proofs, however, Joos's proof for the optimal arrangement of thirteen points in an equilateral triangle requires continuous functions and calculus. I propose that it is possible to continue Melissen's line of reasoning, and complete an entirely discrete proof of Joos's Theorem for the optimal arrangement of thirteen points in an equilateral triangle. In this paper, we make progress towards such a proof. We prove discretely that if either of two points is fixed, Joos's Theorem optimally places the remaining twelve.

KEYWORDS
optimization; packing; equilateral triangle; distance; circles; points; thirteen; maximize

INTRODUCTION

Packing is a class of optimization problems. The objective is to place some number of non-overlapping geometric objects such that they are entirely contained in a larger object leaving as little space remaining in the larger object as possible. Packing problems have an expansive history beginning from when Kepler conjectured the density of the densest ball packing in 1611.[1] In this paper, we consider the problem of packing thirteen congruent circles in an equilateral triangle. It is important to note here the relationship between the problem of packing thirteen congruent circles into an equilateral triangle and maximizing the minimum distance between thirteen points in an equilateral triangle. As **Figure 1** below shows, the centers of the circles in the optimal packing of thirteen congruent circles in the larger triangle are the points that maximize the minimum distance between thirteen points in the smaller triangle. Thus, these are the same problem.

The two interpretations of the problem lead to different applications. By looking at the problem as a packing of circles, we answer questions along the lines of how large a box needs to be in order to hold n bottles of water. By considering the problem in terms of maximizing the distance between points on a plane, we solve problems like how to optimally place transistors onto a microchip. Throughout this paper, we will consider the problem in terms of maximizing the minimum distance between thirteen points in an equilateral triangle.

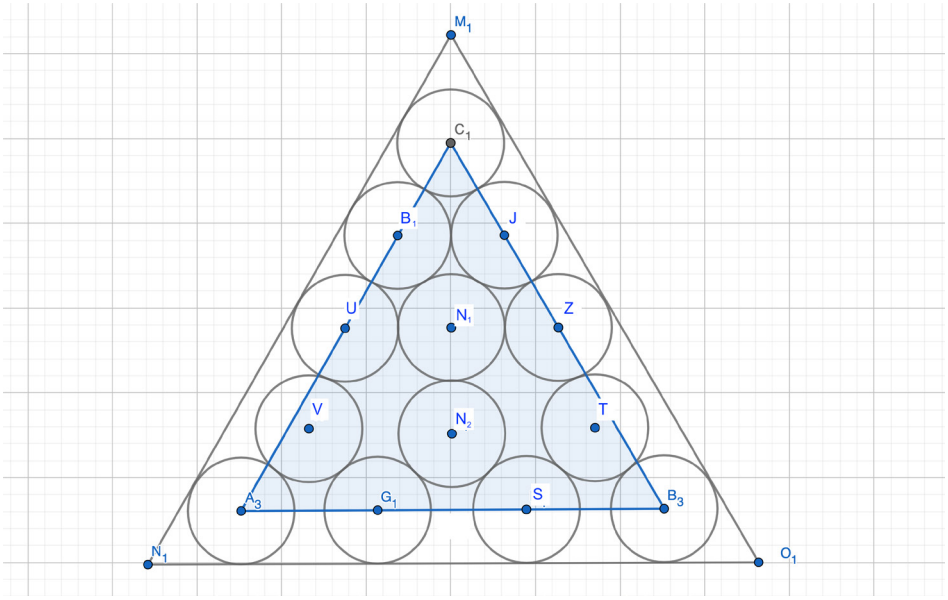

Figure 1. An equilateral triangle with optimal point packing within an equilateral triangle with optimal circle packing

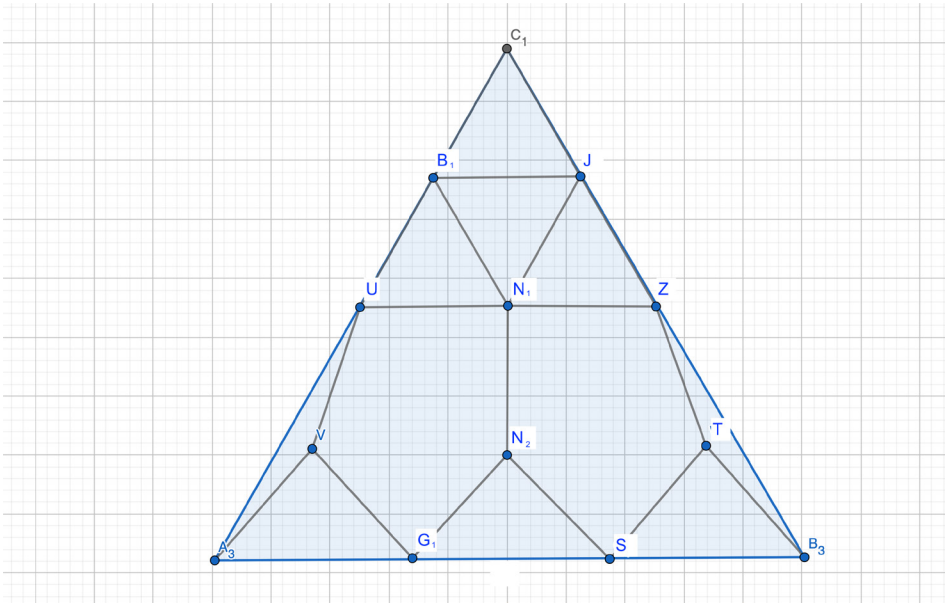

Figure 2. Optimal configuration of thirteen points in an equilateral triangle

Configurations that maximize the minimum distance between n points in an equilateral triangle have been proven for $n \leq 12$,[2, 3] $n = 14, 20$,[4] $n = k(k+1)/2$ for any $k \in N$, [5] and now $n = 13$.[6] $k(k+1)/2$ is the k^{th} triangular number, and the optimal packing is the obvious one: in rows of $1, 2, 3, ..., k$ points. The proof for $n = 13$ was completed by Joos in 2020, confirming Melissen's 1993 conjecture that the orientation of thirteen points in an equilateral triangle shown in **Figure 2** uniquely maximizes the minimum of the distance between two points.[2] In other words, no other configuration of thirteen points in an equilateral triangle exists such that the distance between every pair of points is more than or equal to the minimum distance between the points in **Figure 2**. We will denote this minimum distance as d_{13}.

Joos's proof for the optimal configuration of thirteen points in an equilateral triangle diverged from the strategies that had been used previously to find optimal configurations of points in equilateral triangles. Rather than using only discrete mathematics, his proof requires continuous functions and calculus, relies heavily on inequalities, and considers several cases. This paper explores the possibility of an entirely discrete proof of the optimal arrangement of thirteen

points in an equilateral triangle. We demonstrate that if one can prove that an optimal placement of thirteen points must include N_1 or N_2 from **Figure 2** then we would have an alternate and discrete proof that the Melissen placement of points its optimal. Determining whether the methods that had worked for proving the optimal arrangements of fewer points are still valid for thirteen points could influence whether people must search for new approaches as the number of points continues to increase.

Joos computes $d_{13} = 9 - 5\sqrt{3} - \frac{7\sqrt{6}}{2} + 6\sqrt{2} \approx 0.251813$ for a triangle with side length 1.[6] Pairs of points in **Figure 2** above whose distance is d_{13} are connected by a grey line.

RESULTS

We know from Melissen's 1993 paper that the configuration of points shown in **Figure 2** exists such that pairs of points connected by a grey line in the figure are d_{13} apart.[2] From there, his conjecture states that this configuration is both optimal and unique, meaning that there is no other configuration of thirteen points in an equilateral triangle such that every pair of points is more than or equal to d_{13} apart. We will prove two theorems: one showing that if we assume the position of N_1 from **Figure 2**, then Joos's Theorem for thirteen points holds; and the other showing that if we assume the position of N_2, then Joos's Theorem holds as well.

We first note that Melissen proved the following lemma:

Lemma. In an optimal configuration of $n \geq 3$ points in an equilateral triangle, the three vertices of the triangle must be among the selected points.[3]

Theorem 1. *Let N_1 from **Figure 2** be the point that is d_{13} from each of the points on adjacent sides of the triangle that are d_{13} from the top corner. If N_1 is fixed, then Joos's Theorem for the optimal arrangement of 13 points is correct.*

In other words, given the position of N_1, there is no arrangement of the remaining twelve points in the triangle other than the arrangement shown in **Figure 2** such that every pair of two of the thirteen points is at least d_{13} apart. We first realize that for such a configuration to exist, it cannot contain any other points in the interior of a circle with radius d_{13} centered at N_1. As **Figure 3** shows, this creates two distinct regions: the region above the circle about N_1 and the region below. We will begin by examining the upper region.

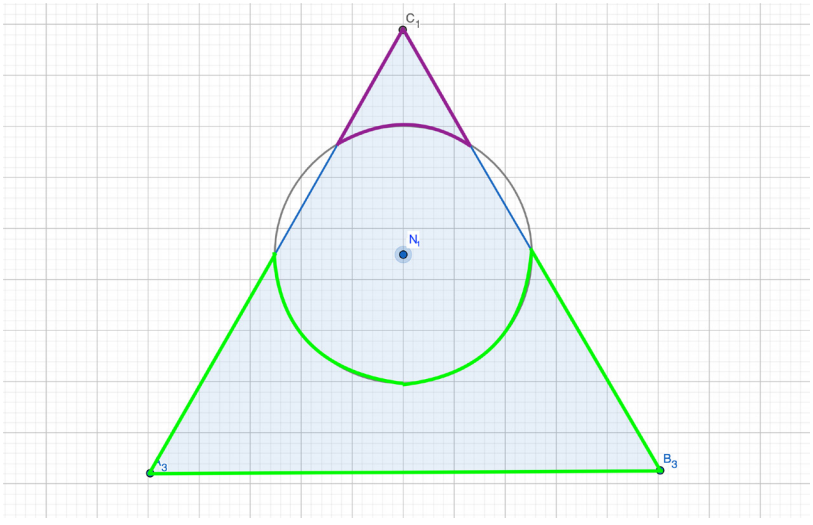

Figure 3. Equilateral triangle with a circle of radius d_{13} about N_1

Given our definition of N_1 and as **Figure 4** illustrates, the intersection of the interiors of circles of radius d_{13} about each of the three corners of the upper region contain every point except the three corners. Thus we know that every point other than the corners is less than d_{13} every other point in the region. So in order for there to be at least two

points in this region whose distance apart is at least d_{13}, those points must be the corners. This also shows that we can include at most three points in this region. We can achieve this only by placing them in the corners, so the arrangement of three points in this region is unique. Since we can fit at most three points in this region, we must put at least nine points in the lower region.

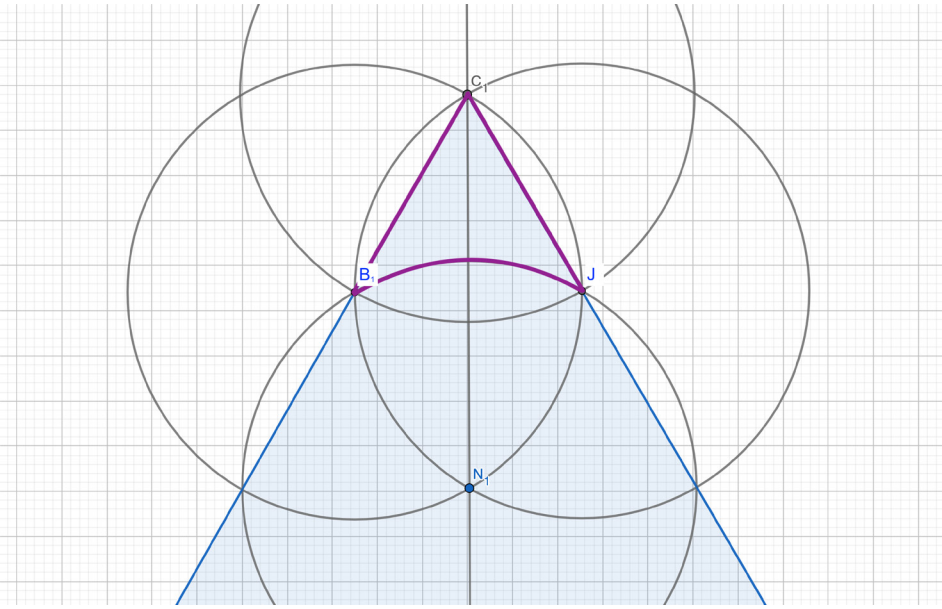

Figure 4. Equilateral triangle with upper region outlined

The decomposition shown in **Figure 5** partitions the lower region into eight subregions: the interiors of circles with radius d_{13} about U and Z (the highest points in the lower region), the interiors of the circles with radius d_{13} about the bottom corners, the interiors of the circles with radius d_{13} about V and T (the intersection points for the first two circles) minus the first two circles, and then the remaining space split symmetrically down the center. The lower region includes its upper boundary, so each of the five arcs that make up this boundary are included in their respective subregions. The purple and orange subregions include their boundaries except where they intersect with brown subregions; these intersections belong to the brown subregions. The brown subregions include their boundaries except where they intersect with the blue subregions. The blue subregions include their entire boundaries.

Since there are nine points, without loss of generality, we must place five left of and including the center. There are five points to place in four subregions. Thus, at least one subregion must have at least two points. Since the circular regions do not include their entire boundaries, they cannot fit more than one point. We take a closer look at the region outlined in blue from **Figure 5** in **Figure 6** below.

Figure 6 shows the region outlined in blue along with a circle of radius d_{13} about each of its corners. The circles about the corners of the region - N_2, H_1, G_1, and M_1 - are green, pink, orange, and brown respectively. Since the pink and brown circles contain the entire region, it is clear that any point in the blue region is less than d_{13} from H_1 and M_1. From Melissen's orginal paper, we know that N_2 (the point d_{13} from and directly below N_1) and G_1 (the point d_{13} from V on the side of the triangle) are d_{13} apart.[2] Thus, the only way to fit two points in this region is at N_2 and G_1.

We have established that the only subregion from **Figure 5** that can fit two points is the blue region and we cannot fit three points in any subregion. So we must put two points in the blue subregion at N_2 and G_1 and one point in each of the other subregions. We know from our lemma that there must be a point at A_3 (the corner). Now the only point

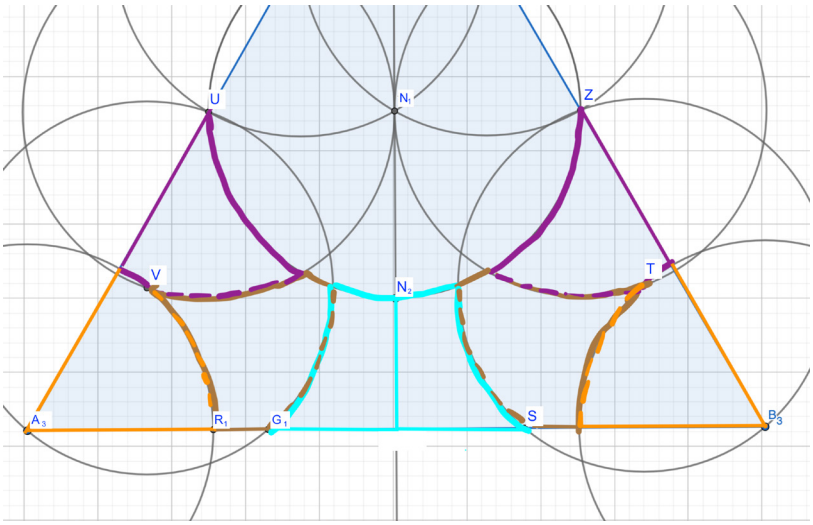

Figure 5. Decomposition of lower region of equilateral triangle

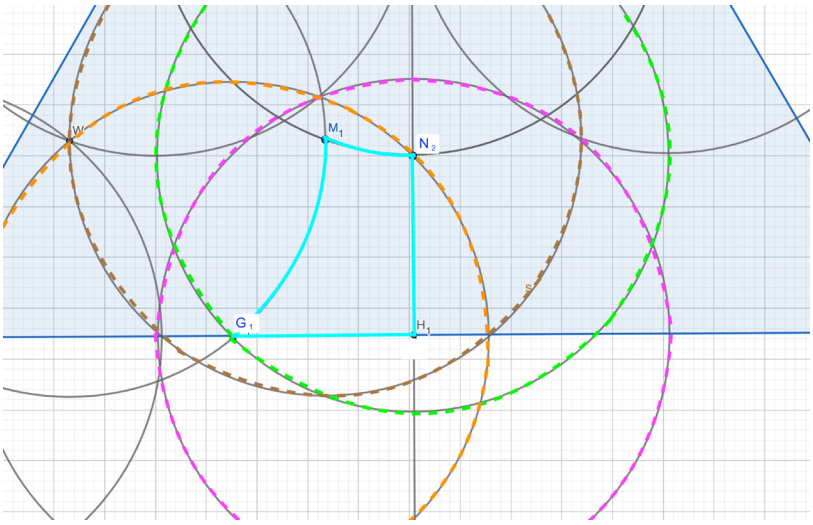

Figure 6. Blue subregion with circles about each corner

in the brown subregion that is d_{13} from G_1 and A_3 is V so we must place the next point there. The only point in the purple region that is d_{13} from V is U, so a point belongs there as well. By the same argument, the points on the right side must be placed in the corresponding locations. We have shown that we can fit at most three points in the upper region and nine points in the lower region and that the arrangements of points in both of those regions is unique, so we have uniquely placed all twelve points. This concludes **Theorem 1**.

Theorem 2. *Let N_2 from **Figure 2** be the point on the altitude from C_1 that is d_{13} from N_1. If N_2 is fixed, then Joos's Theorem for the optimal arrangement of 13 points is correct.*

In order to prove **Theorem 2**, we will first split the triangle into an upper and lower region and decompose the upper region to show that we can fit at most five points in the upper region. We will then decompose the lower region into identical halves, providing that one of these halves must contain four points. We will then show that an arrangement of four points into one of the halves is unique. Fixing these points will further limit the arrangements of points in the upper region, allowing us to change our decomposition and prove that we can actually only fit four points in the upper region, so therefore, there are four points in the other half of the lower region as well. We will then prove that there is a unique configuration of four points in the remaining half of the lower region and four points in the upper region, thus giving a unique configuration of all of the remaining twelve points.

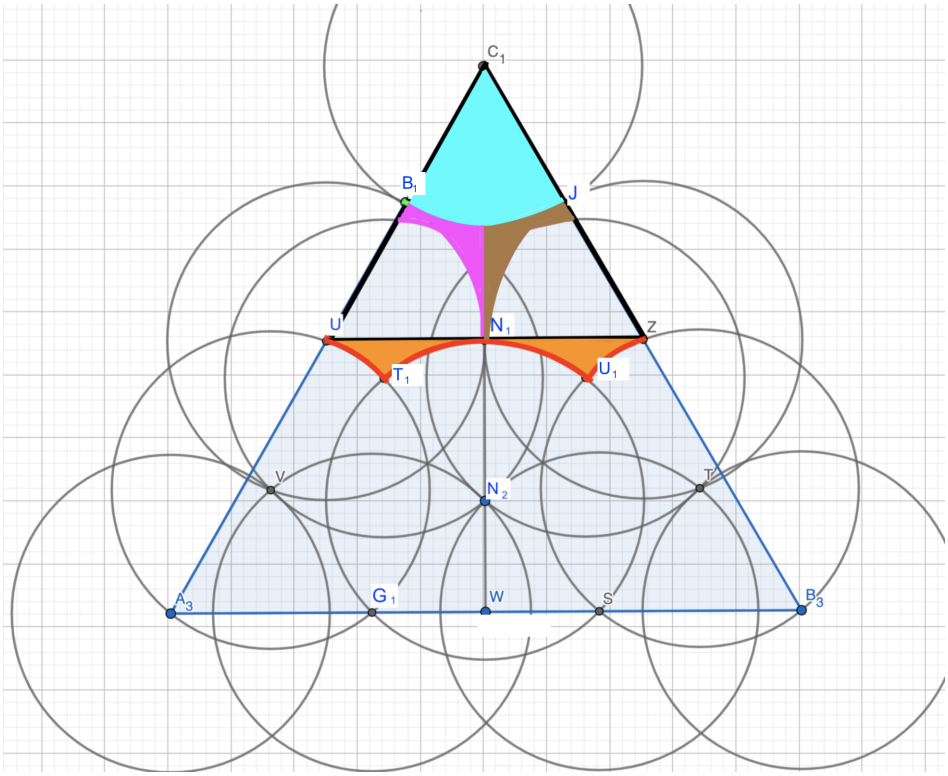

Figure 7. Decomposition of an upper region of an equilateral triangle

In order to maintain at least d_{13} between each point, there cannot be another point within a circle of radius d_{13} centered at N_2. **Figure 7** draws a red curve along the boundary of the union of a circles of radius d_{13} about the points N_2, V and T from **Figure 2**. We will consider the region above and including the red curve - but excluding the points U and Z - as the upper region, and the region below the red curve plus U and Z as the lower region.

Consider equilateral triangle $\triangle C_1UZ$ outlined in black in **Figure 7**. We will first show that there can be at most four points in this triangle excluding U and Z. Each side of this triangle has length $2d_{13}$ since the points are defined such that B_1 is d_{13} from C_1 and U is d_{13} from B_1. By Melissen's proof optimizing six points in an equilateral triangle, we can uniquely fit six points into this triangle such that no pair of points is less than d_{13} apart at C_1, B_1, J, U, Z, and N_1.[2] Since this configuration is unique, if we exclude points U and Z from the region, we can no longer fit six points in the triangle outlined in black. Furthermore, since the optimal configuration of five points in an equilateral triangle requires five points from the configuration of the six points labeled in the region,[2] removing those two points from the region makes it so that we cannot fit five points d_{13} apart from one another either. Thus, in the equilateral triangle outlined in black excluding U and Z, we can fit at most four points.

It follows that in order to fit six points in the upper region, we would need to put two points between the black triangle and the red curve shaded in orange - one on each side of N_1 since we cannot fit two points on either one side given that N_1 is d_{13} from U and Z. We will consider these two orange regions separated by N_1.

In order for there to be a point in each of these regions, there can be no point in the intersection of the circles about N_1, U, and T_1 and the intersection of circles about N_1, U_1, and Z all of radius d_{13}, i. e. there can be no points below the pink and brown regions in **Figure 7**. This is because any point that is in one of those intersections would be less than d_{13} from every point in the corresponding orange region, and we have already established that there must be a point in both regions for there to be six points in the whole upper region.

We next show that we cannot fit four points in the union of the blue, pink, and brown regions. The blue region is defined as the interior of a circle of radius d_{13} about C_1, so there cannot be more than one point in this region. The pink and brown regions can each also have at most one point. To understand why, keep in mind, we took B_1, J, and N_1 from the optimal placement of six points in black equilateral triangle, so we know that they are d_{13} apart. We cannot have a point at N_1 because we know that there must be a point in each of the orange regions and U and Z are excluded from those regions, so each of those regions contains a point less than d_{13} from N_1. Thus, none of the blue, pink, and brown regions can fit more than one point, so we can fit at most three points the union of these three regions. Thus, there are at most five points in the upper region.

We now consider the lower region, which is below the red curve from **Figure 7** and includes U and Z, but excludes the interior of a circle of radius d_{13} about N_2. Since we know we can only fit five points in the upper region, we must fit seven points in this region. Without loss of generality, we must fit four point on the left side of this region.

Figure 8 decomposes the left side of this lower region into three subregions. We define the orange subregion as the interior of a circle of radius d_{13} about U, the green subregion as the interior of a circle of radius d_{13} about A_3 and the brown as the remaining space. We know from Melissen's paper that G_1 and V are d_{13} apart.[2] Thus, the orange and green subregions can each contain at most one point and the brown subregion can contain at most two points uniquely placed at G_1 and V. Since we need to fit four points into these three subregions, we need to fit at least two points in at least one of these subregions, so we must put points at G_1 and V. The only way to put a point in each of the remaining subregions is by putting them at A_3 and U. Thus, the only possible configuration of four points on the left side of this region are G_1, V, A_3, and U.

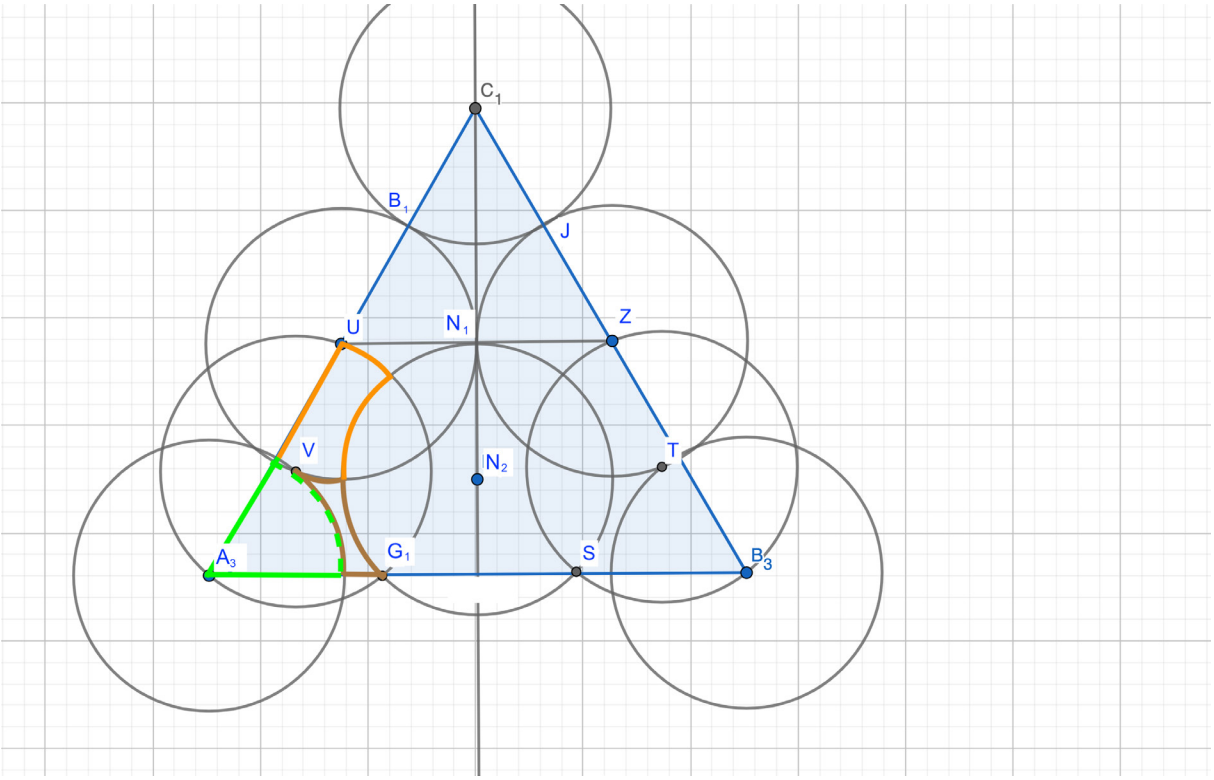

Figure 8. Decomposition of a left lower region of an equilateral triangle

Once we fix those four points, we cannot have any other points within d_{13} of any of them. In particular we cannot have any point in the interior of the circle of radius d_{13} about U which is outlined in pink in **Figure 9**. With this in mind, **Figure 9** revises our decomposition of the upper region using circles of radius d_{13} about C_1 and Z as boundaries of the regions. Note that Z is not in the upper region at all, N_1 is in the red region, J is in the yellow region,

and B_1 is in the blue region. Since each region can contain at most one point, we can now conclude that there are at most four points in this entire upper region. This means that there must also be four points in the right side of the lower region. An analogous decomposition to that shown in **Figure 8** shows that the points on the right side of the lower region must be in analogous locations to those on the left: B_3, S, T, and Z. In particular, there is a point at Z, so the points in the red and yellow subregions in **Figure 9** must be at N_1 and J respectively. Fixing those points also requires that the only points that can be placed in each of the other two regions are B_1 and C_1. This concludes the unique configuration of the remaining twelve points.

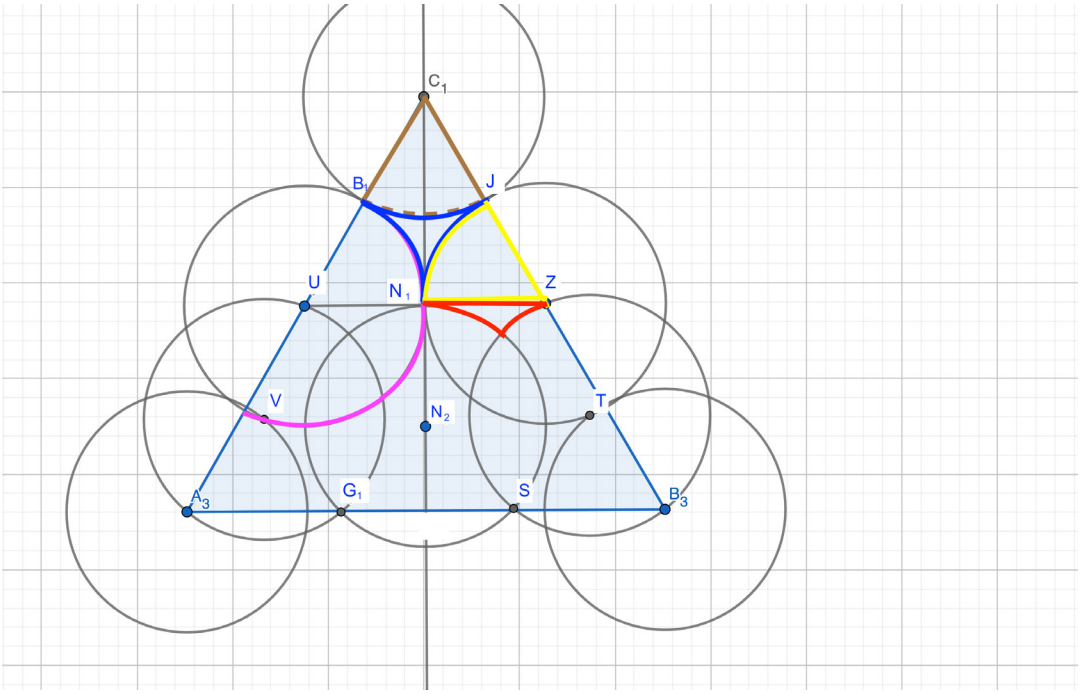

Figure 9. Refined decomposition of the upper region of the equilateral triangle

DISCUSSION

The proofs provided in this paper bring us closer to a discrete proof of Joos's Theorem for the optimal configuration of thirteen points in an equilateral triangle. We have shown that if either of two points is fixed, then we have an alternate proof of the Melissen configuration of thirteen points. We will now explore how one might begin proving the position of N_1 or N_2.

Theorem 3. *Recall N_1 and N_2 from **Figure 2**. Let K, T_1 and W_1, V_1 be in the position that N_1 and N_2 would be in if the entire triangle was rotated about the center such that C_1 was positioned at B_3 and A_3 respectively. Let region H be the orange hexagon from **Figure 10** defined by T_1, N_1, V_1, K, N_2, and W_1 including its boundary. If there are two points in this region at least d_{13} apart, they must be at either N_1 and N_2, K and T_1, or W_1 and V_1 (which by our definition are all the same pair of points up to rotations of the triangle).*

Since H is a convex hexagon and congruent across all three altitudes of the large triangle, proving **Theorem 3** is equivalent to proving that the distances from N_1 to W_1, N_1 to K, and W_1 to K are less than d_{13}.

The point E in **Figure 10** is the midpoint of the segment from C_1 to A_3 which is the side length of the triangle. Assuming the triangle has side length 1, the segment from C_1 to E is 0.5 in length. The segment between C_1 and U from **Figure 2**, on the other hand, has length $2d_{13} \approx 0.503626$. So E from **Figure 10** is closer to B_1 than U from **Figure 2** is. Therefore the distance between E and N_1 is less than d_{13} and equal to the distance between W_1 and E since W_1 is the same distance from A_3 as N_1 is from C_1. Thus, $\triangle EN_1W_1$ is isosceles, so $\angle EN_1W_1 \cong \angle EW_1N_1$. Since $\triangle C_1UZ$ from **Figure 2** is equilateral, $\angle C_1UN_1$ is 60°, so $\angle C_1EN_1$ in **Figure 10** is greater than 60°as is $\angle A_3EW_1$.

Thus, $\angle N_1EW_1$ is less than 60°, and therefore smaller than $\angle VN_1W_1$ and $\angle EW_1N_1$. So, the segment from W_1 to N_1 is less than that from E to N_1 which is less than d_{13}. We can generalize this argument to now say that the segments from N_1 to W_1, N_1 to K, and W_1 to K are less than d_{13} since they are all the same pair of points up to rotations of the triangle. This concludes **Theorem 3**.

Now in order to show that there are points at N_1 and N_2, we need only to show that there must be two points in H because the only way to fit two points in H is by placing points N_1 and N_2, K and T_1, or W_1 and V_1 which are all the same pair of points up to rotations of the triangle.

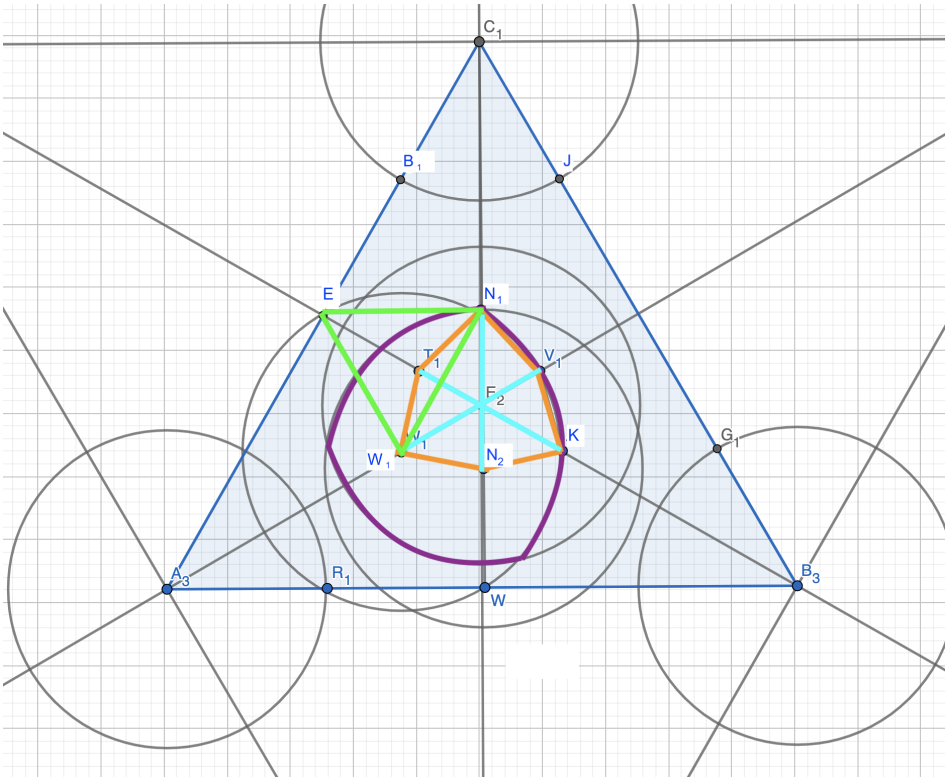

Figure 10. Equilateral triangle with regions of emphasis outlined

In order to show that there are at least two points in H, we would first need to show that there must be at least one point in H, meaning it is impossible to fit thirteen points outside of H that are at least d_{13} apart. We would then need to show that given that there is a point within H, we cannot fit twelve points outside of H that are all at least d_{13} apart. The blue lines clearly partition H into six subregions. Because the six subregions are congruent, once we prove that there is a point in H, we can assume without loss of generality that that point is in the bottom left subregion of H, the triangle defined by the points F_2, W_1, N_2 including the boundary.

If there is a point in that bottom left subregion of H, there cannot be another point that is less than d_{13} from all three corners of the subregion or else it would be less than d_{13} from every point in the subregion. Thus, if there is a point in the bottom left subregion of H, there cannot be any other point within the intersection of the circles of radius d_{13} about F_2, W_1 and N_2 which is outlined in purple (excluding the boundary); we will call this region P. So in order to complete the discrete proof of Joos's Theorem, all that remains to be proven is that we cannot place thirteen points outside of H and that we cannot fit twelve points outside of the union of H and P.

CONCLUSIONS

This paper makes progress towards a discrete proof for the optimal orientation of thirteen points in an equilateral triangle by proving that if either of two points is fixed, the other twelve are positioned in accordance with Joos's Theorem. Finding that such a proof exists could influence how mathematicians approach similar packing problems. It is also interesting to note that 13, along with 8 and 19 can be expressed as $\frac{k(k+1)}{2} - 2$ or $\frac{k(k+1)}{2} + 2k + 1$ depending on the value of k. The first representation shows us that these values are two less than triangular numbers (which are expressed as $\frac{k(k+1)}{2}$), but the second makes sense of the arrangement of the points. Melissen's proof for the optimal configuration of $n = 8$ points, and in his conjectures for $n = 13$ and $n = 19$, all have very similar arrangements. They all have an upper region where the points are arranged as an equilateral triangle with a triangular number of points (that corresponds with the first term of the expression). A discrete proof for $n = 13$ could make leeway for a general proof for all n that are two less than a triangular number.

ACKNOWLEDGEMENTS

The author thanks Professor John Mackey of Carnegie Mellon University for discussions of the problem and reviewing drafts of this manuscript.

REFERENCES

1. Zong, C. (2014) Packing, covering and tiling in two-dimensional spaces, *Expo. Math.* 32, 297–364. *https://doi.org/10.1016/j.exmath.2013.12.002*
2. Melissen, J.B.M. (1993) Densest packings of congruent circles in an equilateral triangle, *Amer. Math. Monthly* 100, 916–925. *https://doi.org/10.2307/2324212*
3. Melissen, J.B.M. (1994) Optimal packings of eleven equal circles in an equilateral triangle, *Acta Math. Hungar* 65, 389–393. *https://doi.org/10.1007/BF01876040*
4. Payan, C. (1997) Empilement de cercles egaux dans un triangle equilateral. A propos d'une conjecture d'Erdos-Oler, *Discrete Math* 165–166, 555–565. *https://doi.org/10.1016/S0012-365X(96)00201-4*
5. Oler, N. (1961) A finite packing problem, *Canad. Math. Bull.* 4, 153–155. *https://doi.org/10.4153/CMB-1961-018-7*
6. Joos, A. (2020) Packing 13 circles in an equilateral triangle, *Aequat. Math.* 35–65. *https://doi.org/10.1007/s00010-020-00753-y*

ABOUT THE STUDENT AUTHOR

Natalie Tedeschi is currently working towards a BS in Discrete Mathematics and Logic at Carnegie Mellon University. She has worked as a teaching assistant in the Department of Computer Science at Carnegie Mellon and as a software engineering intern at Explo, a YCombinator-backed startup. She plans on graduating in May of 2023.

PRESS SUMMARY

How large is the smallest computer chip that contains n transistors that must be some distance apart from one another? How big must a crate be in order to hold n jugs of water? These questions ask how we can maximize the benefit of costly or harmful materials and provide insight on how we can drive technology and innovation forward. Mathematically, these and many more questions are the same. This paper furthers the discussion of solving these types of problems.

An Unbiased Mineral Compositional Analysis Technique for Circumstellar Disks

Yung Kipreos & Inseok Song*

Department of Physics & Astronomy, University of Georgia, Athens, GA

https://doi.org/10.33697/ajur.2021.043

*Student: ykipreos@uga.edu**
Mentor: song@uga.edu

ABSTRACT
A circumstellar disk that surrounds a star is composed of gas, dust, and rocky objects that are in orbit around it. Around infant stars, this disk can act as a source of material that can be used to form planetesimals, which can then accrete more material and form into planets. Studying the mineral composition of these disks can provide insight into the processes that created our solar system. The purpose of this paper is to analyze the mineral composition of these disks by using a newly created python package, Min-CaLM. This package determines the relative mineral abundance within a disk by using a linear regression technique called non-negative least square minimization. The circumstellar disks that are capable of undergoing compositional analysis must have a spectrum with both a detectable mid-infrared excess and prominent silicate features. From our sample, there are only eight debris disks that qualify to be candidates for the Min-CaLM program. The mineral compositions calculated by Min-CaLM are then compared to the Tholen asteroid classification scheme. HD 23514, HD 105234, HD 15407A, BD+20 307, HD 69830, and HD 172555 are found to have a compositions similar to that expected for C-type asteroids, TYC 9410-532-1 resembles the composition of S-type asteroids, and HD 100546 resembles D-type asteroids. Min-CaLM also calculates the mineral compositions of the comets Tempel 1 and Hale-Bopp, and they are used as a comparison between the material in our early solar system and the debris disk compositions.

KEYWORDS
Debris disk; Mineral; Composition; Analysis; Asteroid; Circumstellar; Spectroscopy; Python

INTRODUCTION

Circumstellar disks are rings of gas, dust, and other rocky objects that are in orbit around a star. Two types of common circumstellar disks are protoplanetary disks and debris disks.[1,2] Protoplanetary disks form around infant stars before they reach the main sequence. The rocky material within these disks grows through violent collisions, eventually combining to form planetesimals.[3] Debris disks form after the gas-rich protoplanetary disk around the star has dissipated, and they are composed primarily of dust grains and larger rocky bodies.[4] The material in debris disks is composed of the remnants of collisions between planetesimals around the star. Several dust removal mechanisms eliminate dust grains effectively in the gas-poor debris disks in timescales much shorter than stellar ages, including: (1) Poynting-Robertson drag; (2) stellar wind drag; and (3) stellar radiation blowouts of small grains.[5,6,7] The existence of debris disks indicates that this loss of material must be counterbalanced by the creation of new material via collisions of larger rocky material such as planetesimals.[8] Therefore, the presence of a debris disk suggests the presence of planetesimals or already formed planets around the star.[3]

The light spectrum of debris disks can be used to determine physical characteristics such as the mineral composition, grain sizes, and temperature of the debris disk that emitted it.[9] This paper will primarily study the mineral composition of debris disks. The mineral composition of a debris disk can be determined because the emitted spectrum is a combination of the individual spectra of the dust grains that compose the optically thin disk. Each dust grain mineral species produces a unique spectrum that allows the species to be identified.[10] Thus, the mineral composition of a debris disk can be determined by deconstructing its emitted spectrum into the spectra of the individual mineral species within the disk.

Studying the mineral composition of debris disks in other solar systems may provide insights into the formation of our solar system. A particularly useful analog to the dust in other solar systems is the carbonaceous chondrites that are abundant in the asteroid belt in our solar system.[11] Chondrite meteorites are thought to have been created from shock heating and thermal annealing in the protoplanetary disk of our early solar system.[12] These thermal processes cause them to have a silicate-rich composition that is similar to the mineral composition detected in debris disks.[13] This presents the possibility that the ongoing

processes in the debris disks are the same processes that created our solar system.[14] Therefore, the detection of debris disks with similar compositions to some constituents of the solar system provides a rare opportunity to study the evolution of our solar system.

The challenge of performing mineral spectroscopy on debris disks is the limited number of stellar candidates that meet the necessary qualifications for such analysis. Matthews et al. (2014) reported that debris disks are detected around ~10% of solar-type stars (*e.g.*, the FEPS survey) and around ~20% of FGK type stars (*e.g.*, the DUNES survey).[15,16] However, because of the detection sensitivity required to observe debris disks and the need for a highly sensitive space InfraRed telescope (e.g., Spitzer Space Telescope), the number of currently known stars that have debris disks is relatively small (N=505 from Cotten and Song 2016).[17] The few stars that have debris disks must also have a detectable mid-infrared excess, prominent silicate mineral features, and have been observed by either the Infrared Space Observatory (ISO) or by the Spitzer Space Telescope (SST).[5] For these reasons, out of 505 of the known debris disks, only eight are viable candidates for mineral composition analysis. [17]

Like chondrite meteorites, comets are composed of the material from the early solar system, primarily from the solar nebula during the protoplanetary disk phase.[18] This cometary material is similar in composition to planetesimals found in systems with early planet formation.[19] Therefore, studying the composition of comets is another way to study both the environment of the early solar system, and also the composition of distant planetesimals indirectly.[20] To further study the cometary dust grains in our solar system, NASA conducted the Deep Impact study to measure the spectrum of the comet Tempel 1.[21] In this study, an impactor was sent to collide with Tempel 1.[22] The impact was necessary because the mineralogical emission features required for spectroscopy are only present when sufficiently small grains are abundant.[3] The spectrum of Tempel 1 is representative of the composition of the early solar system when Tempel 1 is thought to have formed.[23] One of the uses of the Deep Impact study is to compare the emission features of the early solar system dust against the emission features of debris disks.[24] For this reason, Tempel 1 is an interesting candidate for mineralogical compositional analysis.

We developed a new mineralogical compositional analysis technique called **Min**eral **C**ompositional **A**nalysis using **L**east Square **M**inimization (Min-CaLM) and this technique was applied to the spectra of various comets and debris disks. We developed the Min-CaLM program to determine the mineral composition and relative abundance of each mineral within it. It should be noted that there are many possible combinations of minerals and relative mineral abundances that can produce the same observed spectrum of a certain target.[11] Ordinarily, mineral compositional analysis is done manually, which could lead to potential bias in the results due to inherent biases the scientist may have. Unlike the traditional method, the Min-CaLM analysis provides a bias-free, purely mathematical composition result. However, we note that Min-CaLM results are not necessarily represent better fits. While it can only be applied to a few debris disks currently, our Min-CaLM method will be useful to analyze many more debris disk spectra that will be obtained with the Next Generation Space Telescope (James Webb Space Telescope).

It must be pointed out that, in this compositional analysis technique, we have made the underlying assumption that each debris disk is optically thin enough that the spectra of the inner disk can be measured. Min-CaLM studies the solid-state features in debris disk spectra, and these features originate in the optically thin region of the debris disks. By making this assumption, the composition calculated by Min-CaLM can be considered to be representative of the composition of the entire debris disk, not just the optically thin regions. This assumption allows the mineral spectra to be added together linearly to recreate the debris disk spectrum. Additionally, it should be noted that Min-CaLM does not consider the effects of different grain temperatures, shapes, or porosities in its calculations. These properties considerably affect the spectra produced by the dust particles, however, in Min-CaLM's mineralogical library, each mineral is represented by a single spectrum with fixed temperature (Black-body), shape (circular), and porosity (minimal porosity).

METHODS

The Debris Disk Selection Process

The debris disks analyzed in this paper were chosen by first examining the list of 505 known debris disks from Cotten and Song (2016) which was the most complete list of nearby (<100 pc) debris disks (> ~5 Myr) debris disks when published. The list did not contain proto-planetary disks because of the age selection criterion. The mid-IR spectra for some of these disks were taken by the IRS spectrometer onboard the Spitzer Space Telescope.[25] These spectra were downloaded from the Spitzer Heritage Archive, which is a library that contains the data from past observations made by the Spitzer telescope.[26] Of the 505 debris disks, Spitzer did not observe 98 of them, and therefore these disks could not be considered for mineral compositional analysis. In order to determine which of the remaining debris disks spectra have strong silicate mineral features, the spectral contributions of the stellar photospheres were subtracted from each of the Spitzer mid-IR debris disk spectra (**Figure 1a**). This subtraction separates the disk spectrum from the stellar photosphere spectrum. Additionally, large dust grains that compose the debris disk produce a smoothly varying blackbody emission that must be removed from the debris disk spectrum before it can undergo mineral spectroscopy.

This approximation of blackbody emission from large dust grains is made so that the silicate spectral features of interest, produced by small dust grains, can be isolated and studied. When there exist strong hints of these dust blackbody contribution, we subtracted dust blackbody contribution with the best-fit blackbody dust temperature obtained by an iterative eye-fit.

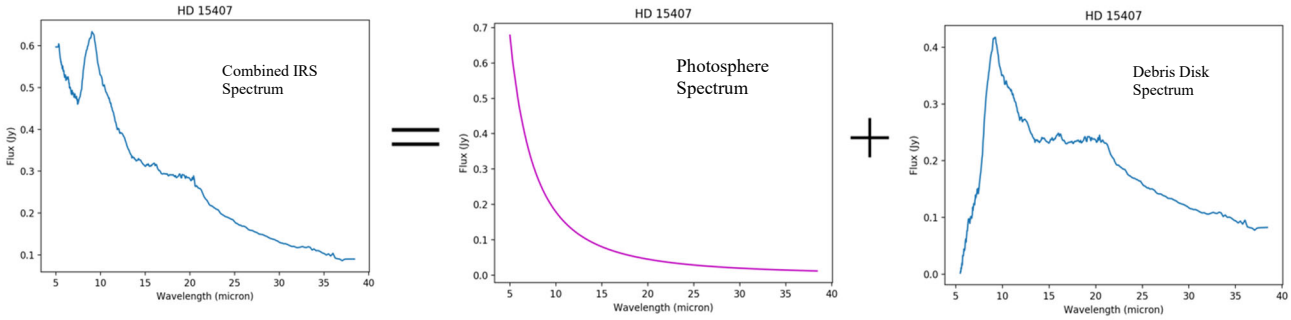

Figure 1. This figure demonstrates that the Spitzer IRS spectrum (left) is a combination of the spectral contributions from the stellar photosphere SED (Spectral Energy Distribution) (middle) and the debris disk spectrum (right). To obtain the isolated debris disk spectrum from the IRS spectrum, the stellar photosphere must be subtracted from the observed IRS spectrum.

Once the debris disk spectrum is isolated and the blackbody component has been removed, the presence of silicate emission features is determined. 397 of the 407 remaining debris disks do not have substantial silicate emission features (**Figure 2**), which leaves only eight candidates that are capable of undergoing mineral compositional analysis. Debris disk spectra with silicate features are rare because for this to be the case, there must be an abundance of small particles in the disk. Particles that are small enough to produce silicate features are easily blown out of the disk via radiation pressure. If a debris disk spectrum has prominent silicate features, then this implies that the small dust grains are being created faster than they are blown away from the disk.[5] The debris disks that this paper will focus on are the disks around the following eight stars: BD+20 307; HD 172555; HD 15407A; HD 23514; HD 69830; TYC9410 532-1; HD 100546; and HD 105234.

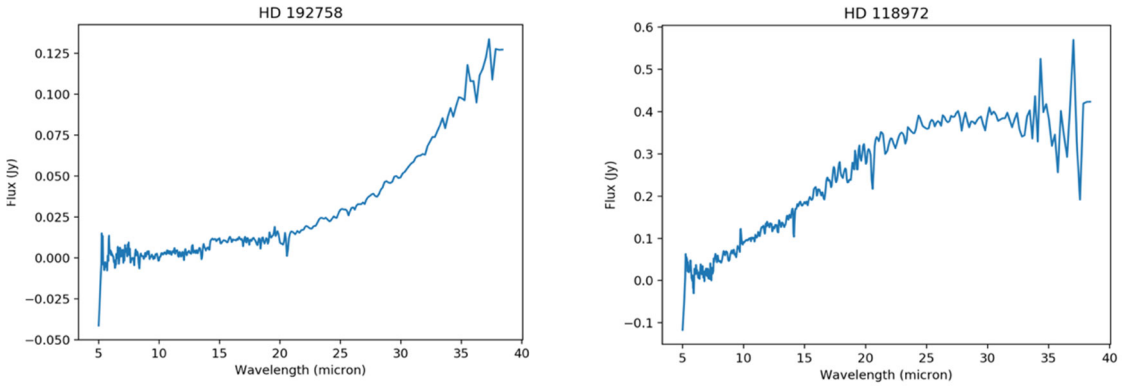

Figure 2. This figure displays two IRS spectra where silicate mineral features are not present. If these spectra contained silicate mineral features, the spectra would resemble the debris disk spectra in **Figure 1**.

In many of the debris disk spectra, there are sections of sharp decreases and increases in flux due to instrument artifacts (**Figure 3**). This issue originates from the reduction of the raw spectra obtained by the Infrared Spectrograph (IRS), an instrument in the Spitzer telescope.[27] There are four spectral modules used in the IRS high resolution observing mode: Short-Low; Short-High; Long-Low; and Long-High. Each module covers a different range of infrared wavelengths. Each of these modules produces a spectrum with a flux scaling that is occasionally offset from the scaling produced by the other modules in the IRS.[28] The spectral peaks from each module section are real features, but they appear to be artificially raised or lowered from the expected flux. One reason for the difference in flux levels is the different widths of the SL and LL slits in each module, which results in different amounts of flux lost.[29] This issue is somewhat mitigated by using the CASSIS library which seeks to calibrate the scaling of each module.[30] The CASSIS database contains optimally extracted spectra measured by each of the different Spitzer IRS modules, unlike the pipeline extracted spectra for general purpose from the Spitzer Heritage Database.[31] For a small fraction of CASSIS reduced-IRS spectra, we see inappropriate inter-module scalings manifested as a sudden discontinuity of spectral shape at the module boundary. To correct the flux scaling for each instrument, we first plot the module spectra together. Misaligned sections

of spectra are multiplied by a scaling factor, which is calculated by dividing the expected flux level by the actual observed flux at that wavelength. As an example, the result of this process on the circumstellar disk surrounding HD 15407A is displayed (**Figure 3**).

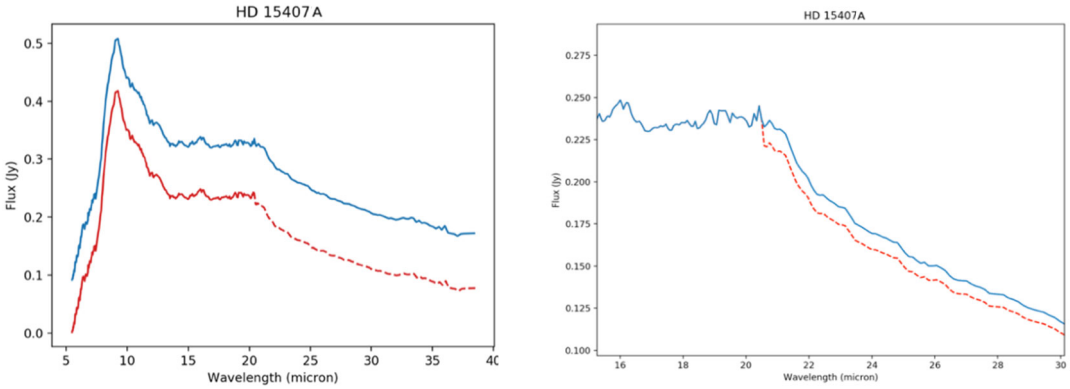

Figure 3. This figure compares the spectrum of the debris disk surrounding the star HD 15407A before (red) and after (blue) the spectrum was flux corrected. In the left figure, both spectra are displayed on top of each other to show the difference between them. The flux corrected spectrum (blue) is raised for demonstration purposes only. In the non-corrected spectrum (red), the flux scaling of the LL1 module (from 21 to 38 microns) is scaled too low compared to the rest of the module flux levels, so a sharp decrease can be seen around 21 microns. The right figure shows a close-up of the discontinuity, with both spectra plotted on top of each other for display purposes. The flux discontinuity can be seen at around ~21 microns. The flux-corrected spectrum should be continuous with no sharp discontinuities, so the LL1 module spectrum must be raised to the correct flux level compared to the other module levels.

Traditional Methods of Mineralogical Analysis

Traditionally, a mineral compositional analysis of debris disks has been performed manually.[24] The minerals that are considered in the mineralogical analysis are chosen based on their likelihood of being present in a debris disk.[23] The minerals are then selected based on the relative strength and wavelength of their spectral features and are then compared to the corresponding features in the debris disk spectrum. The mineral is considered to be a match if its spectral features have a similar height and wavelength to the features in the debris disk spectrum. The relative abundance of a mineral is the amount that its spectrum contributes to the overall debris disk spectrum. Therefore, the relative abundance is related to the amount of that mineral that is present in the disk.[32] The debris disk spectrum is recreated by multiplying each mineral by its relative abundance within the disk and then adding the weighted mineral spectra together. If the minerals and relative abundances are well-chosen, then the newly created spectrum will resemble the original debris disk spectrum.

The Min-CaLM Algorithm

The manually driven phase space search method relies on a human observer to choose a small number of expected mineral species and guide the phase space search by hand, and thus is potentially limited by the initial choice of minerals and bias of the observer and their ability to search phase space. In order to improve this, we began developing the Min-CaLM program to mineralogically fit the IRS spectra of the debris disks around two stars initially, HD 23514 and HD 15407A, and modified the program until it was able to recreate both spectra successfully. The process of this development involved trying different mathematical methods to determine the relative abundance of each mineral to produce the best spectrum fit for both debris disks. The most effective mathematical approach was found to be a linear regression technique called non-negative least square minimization (NNLS).[33] NNLS is a good method for compositional mineral analysis because the target spectrum (debris disk or comet), mineral spectra, and the relative abundances of each mineral create a linear system (**Figure 4**). Additionally, this is an overdetermined system, meaning that there are more equations than there are unknowns. This indicates that the solution to this linear system (*i.e.,* the relative abundances of each mineral) is not unique, and therefore different combinations of mineral abundances could be used to create the same spectrum. NNLS provides a good approximation of the solution of an overdetermined system.[34]

To be able to use NNLS on the debris disk system, first, the target spectrum, mineral spectra, and relative abundances must all be converted into matrix form (**Figure 4**). The mineral spectra matrix is created by placing each mineral spectrum horizontally next to each other. This is a 903×40 size matrix because there are 40 minerals in the Min-CaLM mineral library, and there are 903 spectral points in each mineral spectrum. Likewise, the data points of the debris disk spectrum are resampled into a 903×1 matrix to match the number of spectral points of mineral spectra, and the relative abundances are stored in a 40×1 matrix because each mineral has only one abundance. This system of matrices forms a linear system, as shown in **Figure 4**. In this matrix

representation, $[t_1, t_2, t_3, \dots, t_{903}]$ represents an IRS spectrum of a debris disk, while $[m_{1,i}, m_{2,i}, m_{3,i}, \dots, m_{903,i}]$ represents a mineral spectrum of the i^{th} species in the mineral spectral library, and $[w_1, w_2, w_3, \dots, w_{40}]$ corresponds to the best-fit relative abundances of all mineral species considered.

$$\underset{903 \times 40}{\begin{bmatrix} m_{1,1} & m_{1,2} & m_{1,3} & \cdots & m_{1,40} \\ m_{2,1} & m_{2,2} & m_{2,3} & \cdots & m_{2,40} \\ m_{3,1} & m_{3,2} & m_{3,3} & \cdots & m_{3,40} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{903,1} & m_{903,2} & m_{903,3} & \cdots & m_{903,40} \end{bmatrix}} \underset{40 \times 1}{\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_{40} \end{bmatrix}} = \underset{903 \times 1}{\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ \vdots \\ t_{903} \end{bmatrix}}$$

Figure 4. This is the matrix representation of the linear system containing the mineral spectra (left), the relative abundances of those minerals (middle), and the debris disk spectrum (right). The mineral spectra matrix is of size 903×40. Each column of this matrix corresponds to the individual spectrum of a mineral which contains 903 data points. The relative abundance matrix is of size 40×1 because each mineral has one relative abundance and there are 40 total minerals. The debris disk spectrum of size 903×1 is created by multiplying the mineral spectra and the relative abundance matrices.

Now that each part of the system is in matrix form, the NNLS equation (1) can be applied,

$$argmin||\mathbf{mw} - \mathbf{t}|| \text{ with } w_i \geq 0 \qquad \textbf{Equation 1.}$$

Here **m** represents the mineral spectra matrix, **t** is the target spectrum, and **w** is the relative abundance of each mineral. Additionally, relative abundances are constrained to be zero or positive numbers. The reason for this constraint is that a negative mineral abundance would be non-physical. The result of the solved NNLS equation is a list of 40 mineral abundances. If a mineral is determined by Min-CaLM not to be present in the debris disk, then the abundance of that mineral will be zero. The new debris disk spectrum is recreated by multiplying the spectrum of each mineral with its corresponding relative abundance within the disk. Then each of the weighted spectra is added together. We list all 40 mineral species considered in our Min-CaLM scheme in **Table 1**.

RESULTS

The goal of this study is to calculate the unbiased mineral composition of eight debris disks and two comets using our newly developed mineralogical analysis technique, Min-CaLM. Before using the Min-CaLM scheme, each target spectrum must undergo the photosphere and blackbody removal processes described in the Methods section. The relative mineral abundances calculated by Min-CaLM are displayed in **Table 2** for each target object. These abundances are used to recreate the spectrum of each object, shown in **Figure 5**. Also included in **Table 2** are the black body temperatures used in Min-CaLM's calculations and reduced chi-squared test results for each object. For illustrative purposes, a diagram of the mineral spectra present in the circumstellar disk of BD+20 307 is included in **Figure 7** to demonstrate how the mineral spectra are combined to recreate the observed spectrum.

The reduced chi-square statistic is defined as the chi-square per degree of freedom and is used to determine the goodness of fit of the Min-CaLM recreated and the original spectra. For the reduced chi-square statistic, a fit is considered to be bad if $\chi_\nu^2 > 1$, good if $\chi_\nu^2 < 1$, and over-fitted if $\chi_\nu^2 << 1$.[35] For the purposes of Min-CaLM, an overfitted result is acceptable because the objective of the program is to reproduce the target spectrum as closely as possible. Of the results calculated by Min-CaLM, HD 172555, BD+20 307, HD 100546, and Hale Bopp have chi square results less than one, and HD 15407A, HD 105234, HD 23514, HD 69830, TYC 9410-532-1, and Tempel 1 have chi square results larger than one. It should be noted that any model that produces an acceptable chi-square values should be treated as valid. This system is over-determined, so there are many possible solutions, and if those solutions produce acceptable chi-square results, then they are just as valid as these solutions.

The chi-square test values are calculated with $(obs_i - fit_i)^2 \,/\, noise_i^2$, which is then divided by the degrees of freedom, equal to the number of data points, and where i indicates a wavelength point and the noise is the error recorded by Spitzer. Chi-square values are also plotted against the wavelength for each target object in **Figure 6**. Some diagrams in **Figure 6** have peaks with high chi-square values, but the overall reduced chi-square value is small. This could indicate that the fitting is good at most wavelengths except for the wavelengths with high peaks. Together, these plots illustrate which wavelengths Min-CaLM struggles the most to fit. The chi-square value vs wavelength plot of an optimally reproduced target spectrum would hover around zero, and regions with a raised chi-square value indicate where the Min-CaLM-reproduced spectrum did not correctly reproduce the target

spectrum. For instance, in the plot of HD 100546 in **Figure 6**, Min-CaLM struggled to fit the debris disk spectrum of HD 100546 around 6 and 11 microns, with the largest peaks in the plot at 6.35 microns. This plot indicates that Min-CaLM is possibly missing the species of phyllosilicates and olivine minerals that produce spectral peaks around those wavelengths.[36] As evidenced from the plots in **Figure 6**, the mineralogical library of Min-CaLM is not yet complete and would considerably benefit from the inclusion of more olivine, silicate, and phyllosilicate mineral species.

Mineral Name	Mineral Family	Formula	Description
Tridymite	Tectosilicate	SiO_2	A tectosilicate mineral found in felsic igneous rocks
Albite	Tectosilicate	$NaAlSi_3O_8$	A feldspar mineral found in felsic igneous rocks
Anorthite	Tectosilicate	$NaAlSi_3O_8$	A feldspar mineral found in mafic igneous rocks
Hypersthene	Inosilicate	$(Mg,Fe)SiO_3$	An inosilicate mineral found in igneous rocks and occasionally in metamorphic rocks
Diopside	Inosilicate	$CaMgSi_2O_6$	A silicate mineral found in ultramafic igneous rocks
Hornblende	Inosilicate	$Ca(Mg,Fe)_4Al(Si_7Al)O_{22}(OH,F)_2$	An inosilicate mineral found in metamorphic and igneous rocks
Wollastonite	Inosilicate	$CaSiO_3$	A calcium inosilicate mineral
Nontronite	Phyllosilicate	$Ca_{0.5}(Si_7Al_{0.8}Fe_{0.2})(Fe_{3.5}Al_{0.4}Mg_{0.1})O_{20}(OH)_4$	An iron-rich clay mineral found in basalt rocks
Montmorillonite	Phyllosilicate	$(Na,Ca)_3(Al,Mg)_2Si_4O_{10}(OH)_2(H_2O)$	A member of the smectite mineral group related to clay
Saponite	Phyllosilicate	$Ca_{0.25}(Mg,Fe)_3((Si,Al)_4O_{10})(OH)(H_2O)$	A smectite group mineral found in basalt rocks
Halloysite	Phyllosilicate	$Al_2Si_2O_5(OH)_4$	A clay mineral sometimes found in basalt rocks
Talc	Hydrated Magnesium Silicate	$Mg_3Si_4O_{10}(OH)_2$	A clay mineral commonly found in metamorphic rocks
Olivine	Nesosilicate (Olivine group)	$(Mg,Fe)_2SiO_4$	A magnesium iron silicate commonly found in the upper mantle
Forsterite	Nesosilicate (Olivine group)	(Mg_2SiO_4)	A magnesium-rich olivine mineral commonly found in igneous rocks, metamorphic rocks, and meteorites
Fayalite	Nesosilicate (Olivine group)	Fe_2SiO_4	An iron-rich olivine mineral commonly found in alkaline igneous (specifically volcanic)
Bronzite	Pyroxene	$(Mg,Fe)SiO_3$	Magnesium silicate with containing amounts of iron
Hedenbergite	Pyroxene	$CaFeSi_2O_6$	An iron-rich pyroxene mineral found in chondrites
Diopside	Inosilicate (Pyroxene group)	$MgCaSi_2O_6$	A monoclinic pyroxene mineral commonly found in igneous and mafic rocks
Enstatite	Inosilicate (Pyroxene group)	$MgSiO_3$	A pyroxene silicate mineral commonly found in igneous and metamorphic rocks
Magnesite	Carbonate	$MgCO_3$	A carbonate mineral that contains traces of iron and nickel
Dolomite	Carbonate	$CaMg(CO_3)_2$	An anhydrous carbonate mineral
Siderite	Carbonate	$FeCO_3$	A mineral composed of ~48% iron
Corundum	Oxide	Al_2O_3	A translucent crystalline form of aluminum oxide
Ti3O5 †	Oxide	Ti_3O_5	An oxide mineral
Hematite	Oxide	Fe_2O_3	A gray colored iron ore
Magnetite	Oxide	Fe_2O_4	A magnetic iron ore
SiO *	Other	SiO	Silicon monoxide gas
Quartz	Oxide	SiO_2	A mineral commonly found in felsic igneous rocks
Gamma Alumina †	Oxide	Al_2O_3	A chemical compound related to corundum

Table 1. This table displays the minerals included in the Min-CaLM mineralogical library. Beside each mineral is the mineral group, the chemical formula, and a brief description of the mineral. The mineral group is a collection of minerals based on their chemical compositions. Min-CaLM calculated the emission for 19 of these minerals, but found no traces of them in the debris disk spectra, so those minerals are not displayed in this table. A † symbol means that the mineral spectra was downloaded from the USGS Spectral Library, a * symbol indicates that the spectrum was obtained by tracing a spectral graph, and no symbol means that the mineral spectrum was downloaded from the ASU Spectral Library. The ASU Spectral Library contains mineral absorption spectra, but the Min-CaLM program requires the emission spectrum of each mineral. To convert the mineral absorption spectra to emission spectra, each spectrum is inverted along the y-axis (flux) via the equation [emission flux = 1 – absorption flux] at each wavelength.

The mineral composition of the debris disks is significant because they bear a resemblance to the composition of the asteroids in our solar system. Based on the Tholen Taxonomy classification of asteroids, the compositions of the debris disks calculated by Min-CaLM are similar to the composition of three of the most abundant asteroid types in our solar system. Six debris disks (HD 23514, HD 105234, HD 15407A, BD+20 307, HD 69830, and HD 172555) are found to have compositions similar to that of C-type asteroids, one debris disk (TYC 9410-532-1) resembles the composition of S-type asteroids, and one debris disk (HD 100546) resembles D-type asteroids. These findings will be discussed in further detail in the Discussion section.

Mineral Name	Targets									
	HD 15407A	HD 172555	HD 105234	BD+20 307	HD 23514	HD 100546	HD 69830	TYC 9410 532-1	Tempel 1	Hale Bopp
Blackbody Temperature (K)	344	373	425	424	372	336	330	360	485	-
Chi-square test result	8.75	0.268	15.6	0.00610	59.0	0.0263	126	11.0	606	0.0238
Magnetite	15.1	18.9	-	-	21.7	-	23.6	14.8	24.4	-
Hornblende	15.0	26.9	-	-	9.8	69.0	45.3	12.1	5.5	47.1
Fayalite	-	20.1	-	-	-	16.1	3.5	-	2.8	-
Forsterite	-	-	6.9	8.4	7.4	-	-	10.1	10.1	-
Siderite	0.6	-	3.0	-	2.6	-	-	1.1	3.2	-
Olivine	13.9	15.1	-	8.7	-	7.6	18.3	-	5.8	24.4
Quartz	-	6.2	-	-	6.2	-	-	3.7	0.9	1.2
Tridymite	5.1	-	-	-	19.1	-	-	0.6	4.2	-
Hypersthene	-	-	1.0	33.3	6.3	-	-	-	-	-
Magnesite	-	-	-	-	-	7.3	0.4	1.8	3.2	-
Enstatite	-	-	-	-	-	-	-	-	5.7	11.8
Diopside	-	-	-	3.1	1.4	-	-	5.4	-	-
Dolomite	-	-	-	-	-	-	-	9.6	-	-
Saponite	-	-	49.4	-	-	-	3.8	-	-	-
Talc	-	-	-	18.6	1.7	-	-	8.3	5.0	1.8
Al2O3	6.3	-	5.9	7.6	1.0	-	0.2	8.2	2.4	10.7
Nontronite	-	3.9	-	-	-	-	-	-	-	-
Bronzite	8.1	-	-	-	1.6	-	-	-	-	-
SiO	-	-	12.9	-	-	-	-	-	-	-
Corundum	13.1	-	-	-	8.0	-	5.0	13.4	6.9	2.9
Halloysite	5.6	-	-	7.0	5.2	-	-	2.8	-	-
Albite	2.6	-	-	2.8	4.4	-	-	5.0	5.5	-
Wollastonite	-	-	-	-	1.0	-	-	-	-	-
Anorthite	2.2	-	1.1	0.6	-	-	-	3.3	4.8	-
Gamma Alumina	3.6	-	1.0	10.0	2.7	-	-	-	-	-
Hematite	-	9.0	0.8	-	-	-	-	-	-	-
Ti3O5	-	-	15.0	-	-	-	-	-	6.2	-
Montmorillonite	-	-	3.1	-	-	-	-	-	-	-
Hedenbergite	-	-	-	-	-	-	-	-	3.60	-

Table 2. The relative % abundance of each target spectrum, calculated by the Min-CaLM analysis. There are eight debris disks and two comets. Each target is labeled at the top of the table and the minerals are on the left-hand side. If a mineral is not calculated to be in the target spectrum this is denoted by a "-" symbol. The number in each cell represents the percentage of that mineral calculated to be present in the target spectrum. The spectrum of Hale Bopp does not show a hint of smooth blackbody continuum hence no blackbody temperature in the table.

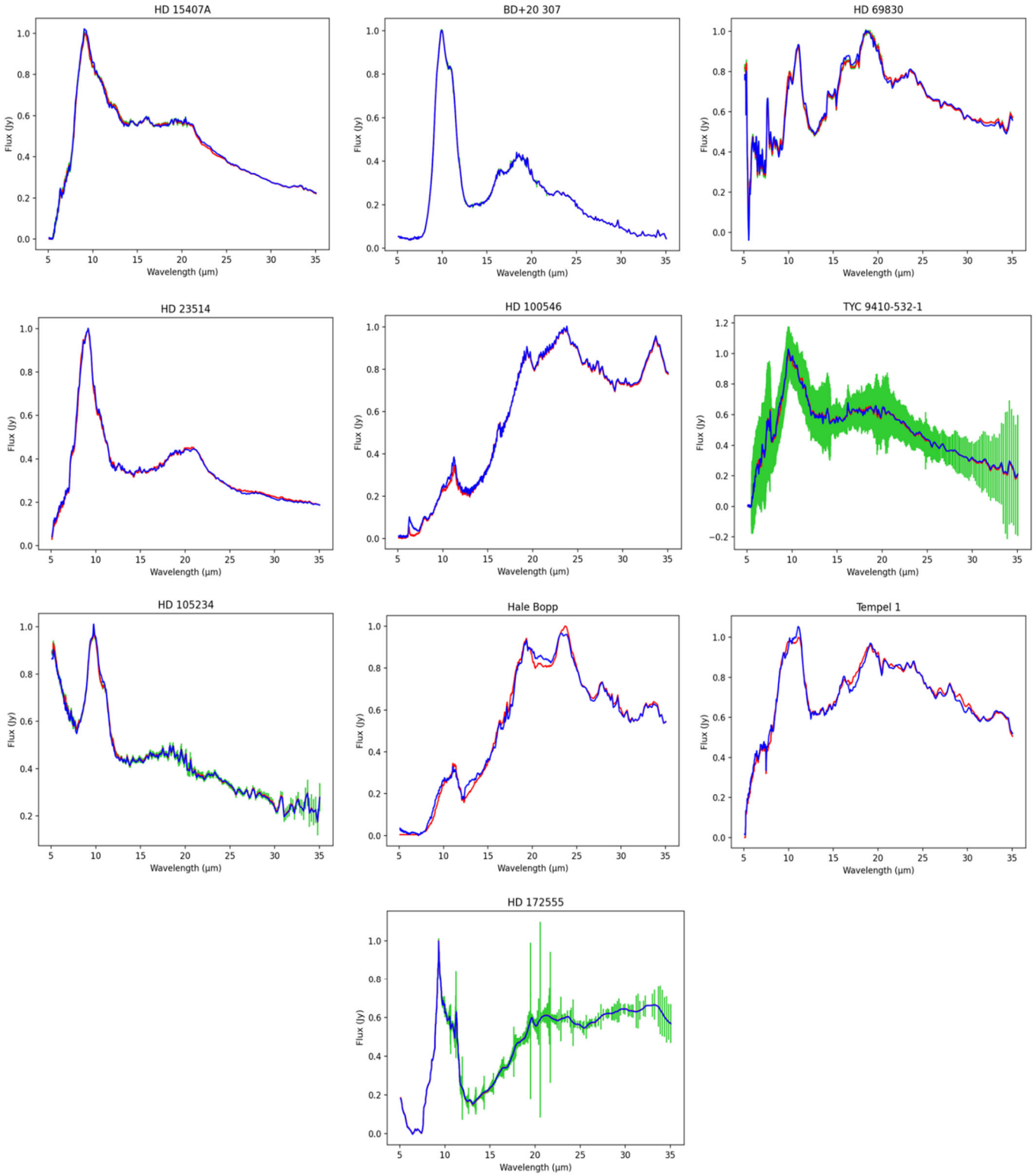

Figure 5. This figure displays the final recreated spectra of the Min-CaLM program. The y-axis of each graph is the flux measured in Jy, and the x-axis is the wavelength measured in μm. The red spectrum is the original photosphere-subtracted disk spectrum of a target object, and the blue spectrum is the recreated best-fit spectrum from Min-CaLM plotted on top of it. The green uncertainty bars represent the uncertainty in the observations made by the Spitzer telescope. It should be noted that in some diagrams, such as the diagram of HD 23514, the Spitzer uncertainty bars from the CASSIS library are too small to be visible. Hale Bopp and Tempel 1 were not observed by Spitzer, and

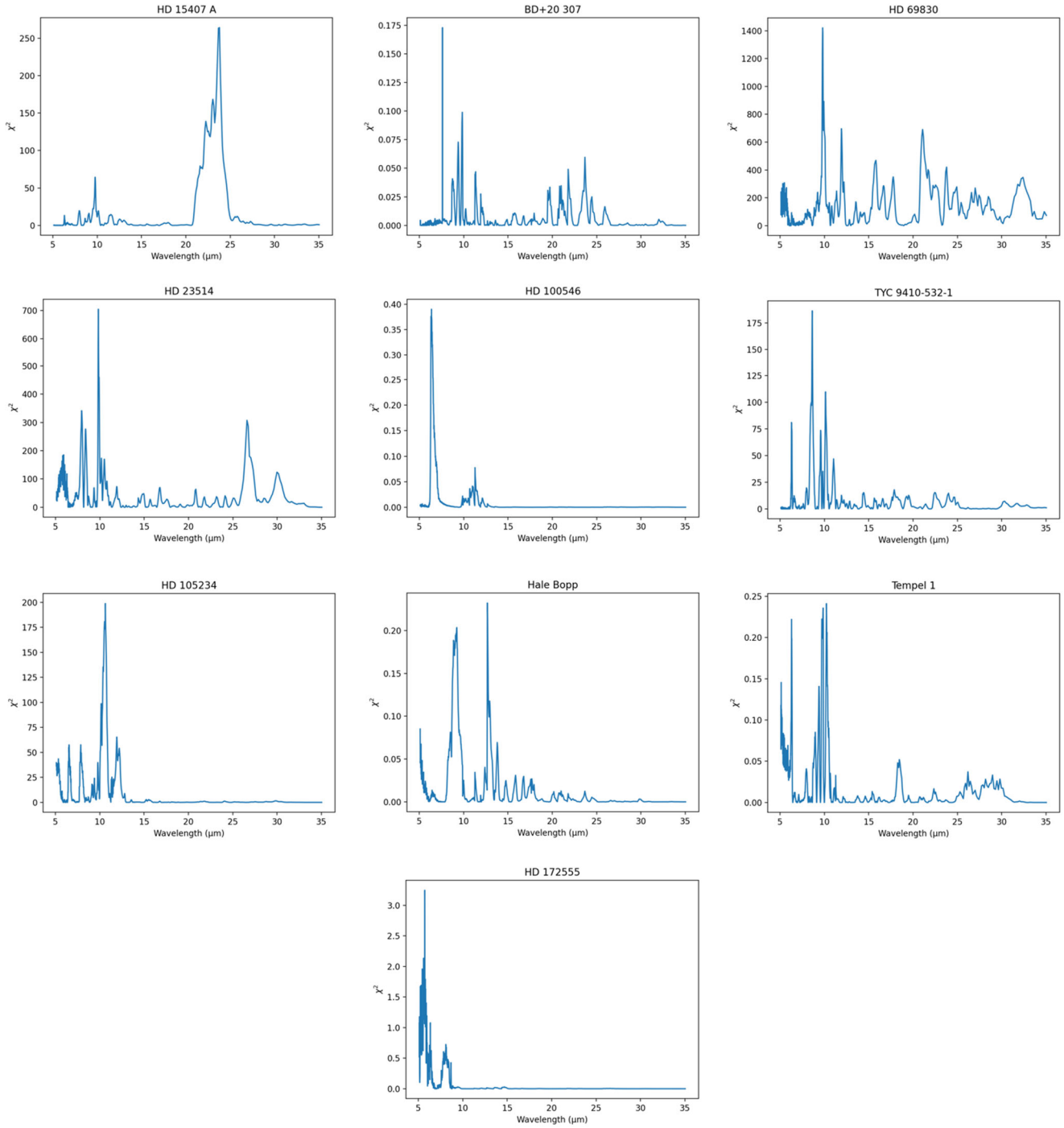

Figure 6. This figure displays the chi-square value vs wavelength for each Min-CaLM calculated spectrum compared to the original measured spectrum. The chi-square value vs wavelength plot for a perfectly matched reproduced spectrum should hover around zero. Therefore, the regions of wavelength that correspond to large chi-square values indicate where Min-CaLM has struggled to properly reproduce the target spectrum. These wavelength regions are thought to have not been properly reproduced because Min-CaLM's mineralogical library is missing the ideal corresponding mineral that contains peaks at that wavelength region.

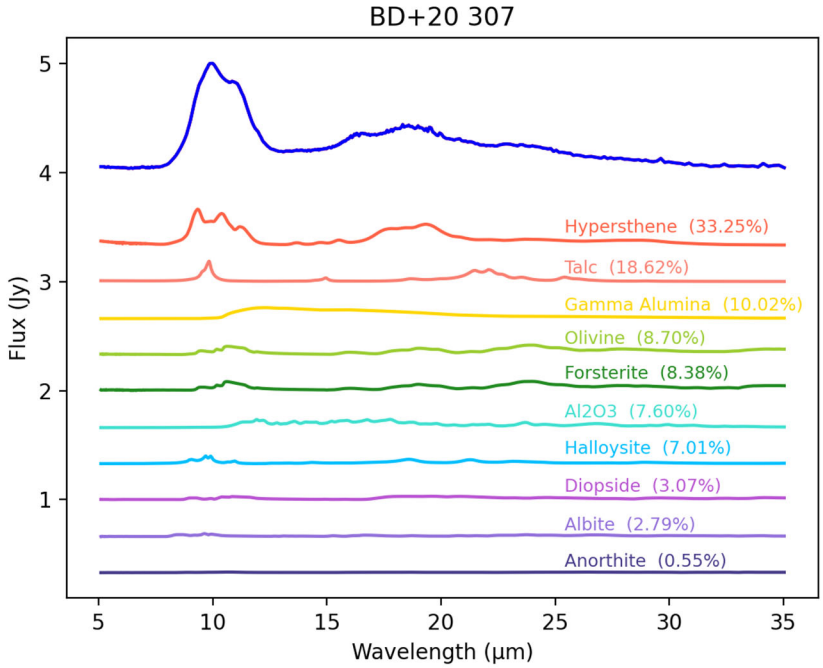

Figure 7. This is a diagram of the Min-CaLM results for the circumstellar disk surrounding BD+20 307. The recreated spectrum is placed at the top (shown in blue) and is overlayed on top of the observed disk spectrum (shown in red). Note that the recreated spectrum (blue) matches the original spectrum (red) to the extent that the original spectrum cannot be seen at this scale. The mineral spectra placed underneath the circumstellar disk spectrum are ordered based on their relative abundances within the disk. All spectra shown in this diagram have been artificially raised and separated for display purposes. The amplitude of the mineral spectra is scaled by their relative abundances in the debris disk.

DISCUSSION

The Relation of Calculated Debris Disk Composition to Our Solar System

Of all known debris disks, around 1–2%, such as the disks analyzed by Min-CaLM, contain a warm infrared excess emission that is created by the collision of planetesimals within the disk.[37] When Min-CaLM calculates the relative mineral abundance of each debris disk, the abundances represent the composition of the planetesimals that created the dust in the collisions. Therefore, a way to relate the Min-CaLM-calculated debris disk compositions to our solar system is to compare them to the asteroids that are present in our system. One way to do this is to use the Tholen Taxonomy of asteroids.

The Tholen Taxonomy of asteroids categorizes asteroids into different classes based on the albedo and spectral shape of each type of asteroid, and each class of asteroids is found in different regions in the solar system.[38] An asteroid class provides insight into the composition of the asteroid and its origin within the solar system.[39] Three of the most abundant asteroid classes found in our solar system are the C, S, and D classes. For comparisons of the composition of the debris disks and asteroid classes, the relative mineral abundances for each debris disk can be found in **Table 2**, and the group of each mineral in the mineralogical library can be found in **Table 1**.

The C asteroid class is believed to be the parent body of carbonaceous chondrite meteorites, meaning that they have a similar composition to chondrites.[40] C-type asteroids are found in the outer region of the asteroid belt and contain a significant amount of hydrated silicates, also known as phyllosilicates.[41] Of the debris disks analyzed by Min-CaLM, six of them, HD 23514, HD 105234, HD 15407A, BD+20 307, HD 69830, and HD 172555 have mineral compositions that resemble the composition that is expected for C-type asteroids. This could indicate that C-type asteroids are abundant in these debris disks because the dust in the disks is likely created by the collisions of such asteroids.

S-type asteroids are found primarily in the inner region of the asteroid belt. The S asteroid class is thought to be the parent body to ordinary chondrites, which are the most common form of chondrites. The composition of S-type asteroids varies widely, and so this class is separated into multiple different subtypes. These subtypes are based on the proportion of olivine minerals to various pyroxene mineral groups, such as clinopyroxenes.[42] Only one of the debris disks analyzed by Min-CaLM is found to have

a similar composition to S-type asteroids. TYC 9410-532-1 resembles the S(III) subtype, which is characterized by a relative abundance of olivine minerals that is much higher than the abundance of minerals in the pyroxene group.[42] For example, the debris disk surrounding TYC 9410-532-1 is calculated to be composed of 10.10% olivine minerals and only 5.40% pyroxene minerals.

The D asteroid class can be found past the outer region of the asteroid belt in our solar system. These asteroids are thought to have originated in the Kuiper belt of our solar system.[43] D-type asteroids are characterized by a lack of phyllosilicates.[44] Of the debris disks analyzed by Min-CaLM, HD 100546 is found not to contain any phyllosilicate minerals and is therefore most represented by the D-type asteroid type.

The spectrum of the gas-rich protoplanetary disk orbiting HD 100546 is known to be remarkably similar to the spectra of the comets Hale-Bopp and Tempel 1.[24] Based on the Min-CaLM best-fit result, these objects consists primarily of oxide and silicate group minerals. The similarity in the composition of the debris disk surrounding HD 100546 and the comets could indicate that the environment in which Hale-Bopp and Tempel 1 were created is similar to the current environment of HD 100546.[45] Because Hale-Bopp and Tempel 1 were likely created in the solar nebula before the creation of planetesimals, this compositional similarity could indicate the presence of forming planetesimals in the circumstellar disk of HD 100546.[24]

Literature Comparison

The mineral compositions calculated by Min-CaLM for circumstellar disks differ slightly from the compositions reported in other papers. This difference is expected because the system of linear equations displayed in **Figure 4** is overdetermined, which indicates that there are many different solutions to the system. The mineral compositions of each debris disk, displayed in **Table 2**, are only one possible set of solutions to this degenerate system. This degeneracy is naturally occurring and cannot be removed from consideration. This becomes more apparent when comparing the composition of debris disks to asteroid types. An example of this can be seen with the composition of HD 69830. The composition that Min-CaLM calculated for HD 69830 is most similar to a C-type asteroid, but in Lisse et al. (2006), it is estimated to have a composition more similar to a P or D asteroid.[24] This difference in asteroid types is a reminder that, because this system is overdetermined, there are many possible solutions for the same system.

Another example of the difference in results can be seen in Lisse et al. (2009). In that paper, the mineral composition of HD 172555 was found to be primarily made of tectosilicates, SiO gas, and olivine minerals. In contrast, Min-CaLM calculated the mineral composition of HD 172555 to contain primarily hornblende, fayalite, and magnetite, which are inosilicates, olivine minerals, and oxide minerals respectively (**Table 2**).

We note that a manual fitting method, such as that applied by Lisse and collaborators, can be beneficial to create traditionally acceptable and appropriate solutions by first considering the minerals that frequently appear in astronomical environments in the fitting procedure. Our method does not take the astrophysical priorities of mineral species into account, and therefore, produces unbiased mathematical solutions. When analyzing a large set of mid-IR disk emission features, such as the ones to be taken with JWST, a quick, unbiased fitting by Min-CaLM followed by detailed manual fitting can be a valuable approach.

Future Improvements

There are several ways that the Min-CaLM program can be improved in the future. A substantial improvement that can be made to Min-CaLM's mineralogical library is to increase the number of minerals that it contains. This expansion of the mineralogical library would include minerals that are found in debris disks, but whose mid-IR spectrum could not be found when creating the library, such as amorphous refractory species. The increase in the number of minerals could potentially make the Min-CaLM-calculated debris disk compositions more scientifically accurate. Currently, the Min-CaLM library contains forty minerals that were chosen by their occurrence in debris disks, as reported in Lisse *et al.* 2006, Lisse *et al.* 2007, and Lisse *et al.* 2009. Our current spectral library of 40 minerals is likely insufficient to fit a variety of debris disk spectra. Because of the missing mineral spectra, some of the spectral features of the debris disks might be unable to be recreated by Min-CaLM. This effect can be seen most clearly in the spectrum recreation of Hale Bopp (**Figure 5**). There is a spectral peak around ~24 μm that Min-CaLM was not able to reproduce, potentially because it is missing the mineral that creates a peak at that wavelength. However, this discrepancy in peak location could also arise from differences in mineral sample grain shape, porosity, annealing history, and crystallinity, all of which Min-CaLM does not take into consideration.[46] Note that an increase of the size of Min-CaLM's mineralogical library would result in an increased number of degenerate solutions, as can be seen in **Table 4**. In this case, the Min-CaLM program would still choose the mineral composition that best fits the debris disk spectrum.

Another issue is that the mineral spectra included in Min-CaLM's mineralogical library only contains spectra from finely milled laboratory samples. When tiny particles are heated, their thermal spectra contain more prominent peaks than the spectrum that is created by larger particles. This problem can be fixed by expanding the mineralogical library to include different grain sizes for each mineral spectrum. With this improvement, the Min-CaLM program output would also include the grain size that was used for each mineral spectrum found in the debris disk.

Besides improvements on the mineralogical library, a potential future change to Min-CaLM could be to add additional linear regression technique such as Non-Negative Weighted Least Square (NNWLS) regression. This technique would place a weighted probability on each mineral based on how likely that mineral is to be found in a debris disk. Additionally, it could be used to make Min-CaLM place greater importance on certain spectral features when recreating the debris disk spectrum. For instance, it might be more crucial to fit the most prominent spectral peaks of the debris disk than it is to fit the ~40 μm tail of the spectrum, which generally contains more noise.[47] Using NNWLS could provide an additional solution to the overdetermined system of the debris disk spectrum, the relative mineral abundances, and the mineral spectra. A caveat to add the NNWLS linear regression technique is that in certain cases, an emphasis on certain spectral features could bias the data analysis, particularly of outlier data.

Another possible improvement to the Min-CaLM algorithm is to use a regression technique called Non-Negative Least Chi-Square regression (NNLC). This regression technique differs from NNLS by minimizing the (chi-square) uncertainty instead of minimizing the residue difference between the observed spectrum and the recreated spectrum, as NNLS does.[48] Within the debris disk spectra studied in this paper, the uncertainty values from IRS spectrophotometric flux are the main source of uncertainty. Taking this into consideration would allow the significance of our results to be quantified. While uncertainty calculations are outside the scope of this paper, in future versions of Min-CaLM this could be remedied by using the NNLC regression technique, or a combination of NNLC and NNWLS.

There are several issues regarding the fitting of the mid-IR spectra of debris disks that this paper and the Min-CaLM program does not take into account. First, the mineral spectra in Min-CaLM's mineralogical library contains mostly pure grains with very little impurities. The "dirtiness" of the dust grains affects how efficiently the grains can absorb energy, and therefore are an important feature to consider. Another issue is that only very small grains produce the mineral features that Min-CaLM studies in debris disk spectra. Any minerals that are bound within large dust grains cannot be studied by Min-CaLM because those grains do not produce silicate mineral features in the mid-IR range. Therefore, the mineral compositions of each debris disk, shown in **Table 2**, are representative of the small grains within the disk, but not necessarily of the large grains. Additionally, the connection between the mineral composition of the dust within the debris disk and the processed material found in asteroids and comets is not entirely straightforward. Small dust grains are blown out of the debris disk by radiation pressure and must be continuously replenished via asteroid or comet collisions for there to be a debris disk. In this paper, we have assumed that small dust grains have a similar composition to the asteroidal material in the debris disk. In nature, dust grains exist with a range of different sizes and a power-law grain size distribution (da/dN ~ a^-3.5) is commonly used to simulate the size effect.[49] Smaller grains are superheated due to their low emissivities and different size grains are all emitting at different temperatures. In Min-CaLM, we are effectively fitting observed spectra with a single grain size emissivity data. This limitation may be the cause of slightly different fitting results obtained in our method versus other literature results. A future improvement is expected to address this grain size issue.

The Min-CaLM program is still in the early stages of development. As previously stated, Min-CaLM does not consider the effects that grain size, different particle temperatures, or porosities in the grains have on the produced spectrum. It also does not consider how the different laboratories produced the input emissivity mineral spectra that are used in Min-CaLM's mineralogical library and what effect that has on the Min-CaLM output. Overall, this paper serves as a demonstration of the potential that this program has in one day becoming a useful tool in the astronomy community. The findings in this paper should be taken with the understanding that the Min-CaLM program is still in development and that future generations of the program will incorporate the improvements discussed in this section.

The Min-CaLM program produces an unbiased mineral composition for debris disks that are capable of undergoing mineralogical compositional analysis. With new debris disk spectral data from the James Webb Space Telescope, the Min-CaLM program and the linear regression method behind it could be refined further to provide a useful platform for future analyses.

Min-CaLM has been published in the Astrophysics Source Code Library (ASCL) under the bibcode 2020ascl.soft01001K. Both the Min-CaLM program and mineralogical library can be downloaded from a GitHub repository that is linked in the ASCL page.

ACKNOWLEDGEMENTS

The authors thank the University of Georgia, the UGA Department of Physics & Astronomy, and the support from the UGA Center for Undergraduate Research Opportunities. The authors thank the referees (two anonymous ones and Dr. Lisse) for their helpful constructive inputs.

REFERENCES

1. Andrews, S., Wilner, D., Hughes, A., Qi, C., Dullemond, C. (2009) "PROTOPLANETARY DISK STRUCTURES IN OPHIUCHUS Protoplanetary Disk Structures in Ophiuchus", *The Astrophysical Journal*, 700(2), 1502–1523. *doi:10.1088/0004-637X/700/2/1502*

2. Henning, T. (2010). Cosmic Silicates. Annual Review of Astronomy and Astrophysics, 48(1), pp.21–46. *https://doi.org/10.1146/annurev-astro-081309-130815*

3. Takasawa, S., Nakamura, A., Kadono, T., Arakawa, M., Dohi, K., Ohno, S., Seto, Y., Maeda, M., Shigemori, K., Hironaka, Y., Sakaiya, T., Fujioka, S., Sano, T., Otani, K., Watari, T., Sangen, K., Setoh, M., Machii, N., Takeuchi, T. (2011) "Silicate Dust Size Distribution from Hypervelocity Collisions: Implications for Dust Production in Debris Disks", *The Astrophysical Journal*, 733(2), L39. *https://doi.org/10.1088/2041-8205/733/2/L39*

4. Su, K., Rieke, G., Stansberry, J., Bryden, G., Stapelfeldt, K., Trilling, D., Muzerolle, J., Beichman, C., Moro-Martin, A., Hines, D., Werner, M. (2006) "Debris Disk Evolution around A Stars", *The Astrophysical Journal*, 653(1), 675–689. *https://doi.org/10.1086/508649*

5. Hughes, A., Duchêne, G. and Matthews, B. (2018). Debris Disks: Structure, Composition, and Variability. *Annual Review of Astronomy and Astrophysics*, 56(1), pp.541–591. *https://doi.org/10.1146/annurev-astro-081817-052035*

6. Plavchan, P., Jura, M. and Lipscy, S. (2005). "Where Are the M Dwarf Disks Older Than 10 Million Years?". *The Astrophysical Journal*, 631(2), pp.1161–1169. *https://doi.org/10.1086/432568*

7. Klahr, H., Lin, D. (2001) "Dust Distribution in Gas Disks: A Model for the Ring around HR 4796A", *The Astrophysical Journal*, 554(2), 1095–1109. *https://doi.org/10.1086/321419*

8. Wood, K., Lada, C., Bjorkman, J., Kenyon, S., Whitney, B., Wolff, M. (2002) "Infrared Signatures of Protoplanetary Disk Evolution", *The Astrophysical Journal*, 567(2), 1183–1191. *https://doi.org/10.1086/338662*

9. Wolf, S., Hillenbrand, L. (2003) "Model Spectral Energy Distributions of Circumstellar Debris Disks. I. Analytic Disk Density Distributions", *The Astrophysical Journal*, 596(1), 603-620. *https://doi.org/10.1086/377638*

10. Clark, R. (1999) Remote Sensing for the Earth Sciences: Manual of Remote Sensing, 3rd ed, John Wiley & Sons, Inc.

11. Morlok, A., Koike, C., Tomioka, N., Mann, I., Tomeoka, K. (2010) "Mid-infrared spectra of the shocked Murchison CM chondrite: Comparison with astronomical observations of dust in debris disks", *Icarus*, 207(1), 45–53. *https://doi.org/10.1016/j.icarus.2009.11.018*

12. Scott, E., Krot, A. (2005) "Thermal Processing of Silicate Dust in the Solar Nebula: Clues from Primitive Chondrite Matrices", *The Astrophysical Journal*, 623(1), 571–578. *https://doi.org/10.1086/428606*

13. Hewins, R., Zanda, B., Bendersky, C. (2012) "Evaporation and recondensation of sodium in Semarkona Type II chondrules", *Geochimica et Cosmochimica Acta*, 78, 1-17. *https://doi.org/10.1016/j.gca.2011.11.027*

14. Moro-Martín, A. (2007) "On the solar system—debris disk connection", *Proceedings of the International Astronomical Union*, 3(S249), 347–354. *https://doi.org/10.1017/S1743921308016803*

15. Meyer, M.R., Hillenbrand, L.A., Backman, D., Beckwith, S., Bouwman, J., Brooke, T., Carpenter, J., Cohen, M., Cortes, S., Crockett, N., Gorti, U., Henning, T., Hines, D., Hollenbach, D., Kim, J.S., Lunine, J., Malhotra, R., Mamajek, E., Metchev, S., Moro-Martin, A., Morris, P., Najita, J., Padgett, D., Pascucci, I., Rodmann, J., Schlingman, W., Silverstone, M., Soderblom, D., Stauffer, J., Stobie, E., Strom, S., Watson, D., Weidenschilling, S., Wolf, S., Young, E. (2006) "The Formation and Evolution of Planetary Systems: Placing Our Solar System in Context with Spitzer," *Publications of the Astronomical Society of the Pacific*, 118(850), 1690–1710.

16. Eiroa, C., Marshall, J.P., Mora, A., Montesinos, B., Absil, O., Augereau, J.C., Bayo, A., Bryden, G., Danchi, W., del Burgo, C., Ertel, S., Fridlund, M., Heras, A.M., Krivov, A.V., Launhardt, R., Liseau, R., Löhne, T., Maldonado, J., Pilbratt, G.L., Roberge, A., Rodmann, J., Sanz-Forcada, J., Solano, E., Stapelfeldt, K., Thébault, P., Wolf, S., Ardila, D., Arévalo, M., Beichmann, C., Faramaz, V., González-García, B.M., Gutiérrez, R., Lebreton, J., Martínez-Arnáiz, R., Meeus, G., Montez, D., Olofsson, G., Su, K.Y.L., White, G.J., Barrado, D., Fukagawa, M., Grün, E., Kamp, I., Lorente, R., Morbidelli, A., Müller, S., Mutschke, H., Nakagawa, T., Ribas, I., Walker, H. (2013) "DUst around NEarby Stars. The survey observational results," *Astronomy & Astrophysics*, 555. *https://doi.org/10.1051/0004-6361/201321050*.

17. Cotten, T., Song, I. (2016) "A Comprehensive Census of Nearby Infrared Excess Stars", *The Astrophysical Journal Supplement Series*, 225(1), 15. *https://doi.org/10.3847/0067-0049/225/1/15*

18. Mumma, M. (1997) "Organic Volatiles in Comets: Their Relation to Interstellar Ices and Solar Nebula Material", *From Stardust to Planetesimals*, 122, 369.

19. Huebner, W. (2002) "Composition of Comets: Observations and Models", *Earth, Moon, and Planets*, 89(1–4), 179–195. *https://doi.org/10.1023/A:10215068*

20. Wooden, D., Harker, D., Woodward, C. (1999) "Crystalline Silicates, Comets, and Protoplanetary Disk Evolution", Thermal Emission Spectroscopy and Analysis of Dust, Disks, and Regoliths, Proceedings of a meeting held at at The Lunar and Planetary Institute, vol. 196(Astronomical Society of the Pacific Conference Series), 99–108.

21. Harker, D., Woodward, C., Wooden, D. (2005) "The Dust Grains from 9P/Tempel 1 Before and After the Encounter with Deep Impact", *Science*, 310(5746), 278–280.

22. A'Hearn, M., Belton, M., Delamere, W., Thomas, P., Kissel, J., Klaasen, K., McFadden, L.,Meech,M.J, Melosh, H.J., Shultz, P.H., Sunshine, J.M., Veverka, J., Yeomans, D.K., Baca, M.W., Busko, I., Crokett, C.J., Collins, S.M., Desnoyer, M., Eberhardy, C.A., Ernst, C.M., Farnham, T.L., Feaga, L., Groussin, O., Hampton, D., Ipatov, S.I., Li, J.Y., Linder, D., Lisse, C.M., Mastrodemos, N., Owen Jr., W.M., Richardson, J.E., Wellmitz, White, R.L. (2005) "Deep Impact: Excavating Comet Tempel 1", *Science*, 310(5746), 258–264. *https://doi.org/10.1126/science.1118923*

23. Lisse, C., VanCleve, J., Adams, A., A'Hearn, M., Fernández, Y., Farnham, T., Armus, L., Grillmair, C., Ingalls, J., Belton, M., Groussin, O., McFadden, L., Meech, K., Schultz, P., Clark, B., Feaga, L., Sunshine, J. (2006) "Spitzer Spectral Observations of the Deep Impact Ejecta", *Science*, 313(5787), 635–640. *https://doi.org 10.1126/science.1124694*

24. Lisse, C., Beichman, C., Bryden, G., Wyatt, M. (2007) "On the Nature of the Dust in the Debris Disk around HD 69830", *The Astrophysical Journal*, 658(1), 584–592. *https://doi.org/10.1086/511001*

25. Backman, D., Marengo, M., Stapelfeldt, K., Su, K., Wilner, D., Dowell, C., Watson, D., Stansberry, J., Rieke, G., Megeath, T., Fazio, G., Werner, M. (2008) " Epsilon Eridani's Planetary Debris Disk: Structure and Dynamics Based on Spitzer and Caltech Submillimeter Observatory Observations", *The Astrophysical Journal*, 690(2), 1522–1538. *https://doi.org/10.1088/0004-637X/690/2/1522*

26. Wu, X., Roby, T., Ly, L. (2010) "Spitzer Heritage Archive", *Observatory Operations: Strategies, Processes, and Systems. https://doi.org/10.1117/12.857728*.

27. Brandl, B., Bernard-Salas, J., Spoon, H., Devost, D., Sloan, G., Guilles, S., Wu, Y., Houck, J., Weedman, D., Armus, L., Appleton, P., Soifer, B., Charmandaris, V., Hao, L., Higdon, J., Herter, T. (2006) "The Mid-Infrared Properties of Starburst Galaxies from Spitzer-IRS Spectroscopy", *The Astrophysical Journal*, 653(2), 1129–1144.

28. Wu, Y., Charmandaris, V., Hao, L., Brandl, B., Bernard-Salas, J., Spoon, H., Houck, J. (2006) "Mid-Infrared Properties of Low-Metallicity Blue Compact Dwarf Galaxies from the Spitzer Infrared Spectrograph", *The Astrophysical Journal*, 639(1), 157–172. *https://doi.org/10.1086/499226*.

29. Hernán-Caballero, A., Spoon, H., Lebouteiller, V., Rupke, D. and Barry, D. (2015). The infrared database of extragalactic observables from Spitzer – I. The redshift catalogue. *Monthly Notices of the Royal Astronomical Society*, 455(2), pp.1796–1806. *https://doi.org/10.1093/mnras/stv2464*

30. Combined Atlas Of Sources With Spitzer IRS Spectra: Search [online] (2019) [online], *cassis.sirtf.com*.

31. Lebouteiller, V., Barry, D., Goes, C., Sloan, G., Spoon, H., Weedman, D., Bernard-Salas, J. and Houck, J., 2015. "CASSIS: The Cornell Atlas of Spitzer /Infrared Spectrograph Sources. II. High-resolution Observations". *The Astrophysical Journal Supplement Series*, 218(2), p.21. *https://doi.org/10.1088/0067-0049/218/2/21*.

32. Currie, T., Lisse, C., Sicilia-Aguilar, A., Rieke, G., Su, K. (2011) " Spitzer Infrared Spectrograph Spectroscopy of the 10 Myr Old EF Cha Debris Disk: Evidence for Phyllosilicate-Rich Dust in the Terrestrial Zone", *The Astrophysical Journal*, 734(2), 115. *https://doi.org/10.1088/0004-637X/734/2/115*

33. Zyskind, G. (1967) "On Canonical Forms, Non-Negative Covariance Matrices and Best and Simple Least Squares Linear Estimators in Linear Models", *The Annals of Mathematical Statistics*, 38(4), 1092–1109. *https://doi.org/10.1214/aoms/1177698779*

34. Williams, G. (1990) "Overdetermined Systems of Linear Equations", *The American Mathematical Monthly*, 97(6), 511. *https://doi.org/10.1080/00029890.1990.11995638*

35. Andrae, R., Schulze-Hartung, T., Melchior, P. (2020) Dos and Don'ts of Reduced Chi-Squared [online], *Arxiv.org*, available: https://arxiv.org/pdf/1012.3754.pdf [accessed 25 Oct 2020].

36. Morlok, A., Lisse, C., Mason, A., Bullock, E., Grady, M. (2014) "Mid-infrared spectroscopy of components in chondrites: Search for processed materials in young Solar Systems and comets", *Icarus*, 231, 338-355. *https://doi.org/10.1016/j.icarus.2013.12.018*.

37. Morlok, A., Mason, A., Anand, M., Lisse, C., Bullock, E., Grady, M. (2014) "Dust from collisions: A way to probe the composition of exo-planets?", *Icarus*, 239, 1-14. *https://doi.org/10.1016/j.icarus.2014.05.024*.

38. Lodders, K., Fegley, B. (1998) *The Planetary Scientist's Companion*, Oxford University Press: New York.

39. Britt, D., Tholen, D., Bell, J., Pieters, C. (1992) "Comparison of asteroid and meteorite spectra: Classification by principal component analysis", *Icarus*, 99(1), 153–166. *https://doi.org/10.1016/0019-1035(92)90179-B*

40. Busarev, V. (2012) "A hypothesis on the origin of C-type asteroids and carbonaceous chondrites", *Asteroids, Comets, Meteors*, 1667.

41. Richard, N. (2002) "The Cambridge encyclopedia of meteorites", *Choice Reviews Online*, 40(02), 40–0661–40–0661.

42. Gaffey, M., Burbine, T., Binzel, R. (1993) "Asteroid spectroscopy: Progress and perspectives", *Meteoritics*, 28(2), 161–187. *https://doi.org/10.1111/j.1945-5100.1993.tb00755.x*

43. McKinnon, W. (2008) "On The Possibility Of Large KBOs Being Injected Into The Outer Asteroid Belt", *American Astronomical Society*, Vol. 40, 464.
44. Cruikshank, D., Dalle Ore, C., Roush, T., Geballe, T., Owen, T., de Bergh, C., Cash, M., Hartmann, W. (2001) "Constraints on the Composition of Trojan Asteroid 624 Hektor", *Icarus*, 153(2), 348–360.
45. Bouwman, J., de Koter, A., Dominik, C., Waters, L. (2003) "The origin of crystalline silicates in the Herbig Be star HD 100546 and in comet Hale-Bopp", *Astronomy & Astrophysics*, 401(2), 577–592. *https://doi.org/10.1051/0004-6361:20030043*
46. Olofsson, J., Juhász, A., Henning, T., Mutschke, H., Tamanai, A., Moór, A., Ábrahám, P. (2012) "Transient dust in warm debris disks", *Astronomy & Astrophysics*, 542, A90. *https://doi.org/10.1051/0004-6361/201118735.*
47. Teplitz, H., Desai, V., Armus, L., Chary, R., Marshall, J., Colbert, J., Frayer, D., Pope, A., Blain, A., Spoon, H., Charmandaris, V., Scott, D. (2007) "Measuring PAH Emission in Ultradeep Spitzer IRS Spectroscopy of High-Redshift IR-Luminous Galaxies", *The Astrophysical Journal*, 659(2), 941–949. *https://doi.org/10.1086/512802*
48. Désesquelles, P., Ha, T.M.H., Korichi, A., Blanc, F.L., Petrache, C.M. (2009) "NNLC: non-negative least chi-square minimization and application to HPGe detectors," *Journal of Physics G: Nuclear and Particle Physics*, 36(3), 037001. *https://doi.org/10.1088/0954-3899/36/3/037001.*
49. MacGregor, Meredith A., et al. "Constraints on Planetesimal Collision Models in Debris Disks." *The Astrophysical Journal*, vol. 823, no. 2, 2016, p. 79., doi:10.3847/0004-637x/823/2/79. *http://dx.doi.org/10.3847/0004-637X/823/2/79*

ABOUT STUDENT AUTHORS

Yung Kipreos graduated from the University of Georgia in May 2020 with a major in Astrophysics and a minor in Computer Science. She wants to attend graduate school to become an exoplanetary scientist.

PRESS SUMMARY

A circumstellar disk surrounding a star is composed of gas, dust, and rocky objects in orbit around the star. Around infant stars, this disk can act as a source of material to form planetesimals, which can then accrete more material and form into planets. Studying the mineral composition of these disks can provide insight into the processes that created our solar system. In this paper, we introduce a new unbiased mineral compositional analysis technique, Min-CaLM, and apply it to eight circumstellar disks around the stars: HD 23514, HD 105234, HD 15407A, BD+20 307, HD 69830, and HD 172555. The Min-CaLM python package is open-access and is available to download from https://github.com/yungkipreos/Min-CaLM.

Do Warmups Predict Pole Vault Competition Performance?

*Alex Peskin**

Honors College, Grand Valley State University, Allendale, MI

https://doi.org/10.33697/ajur.2021.044

*Student: peskinal@mail.gvsu.edu**
Mentor: deanerr@gvsu.edu

ABSTRACT
The aim of this research was to determine the relationship between pole vault warmup and competition performance in a sample of 16 collegiate vaulters over 60 observations. Pole vault athletes are given time to warm up in the same area that the competition will take place. This prompted investigation into whether better warmup performance could indicate better familiarity with the performance environment, and whether this could translate to the competition. The number of warmup vaults taken was also considered. Participants were observed during multiple warmup periods and data was collected on warmup performance. The findings indicate a significant correlation between instances in which participants displayed their best warmup scores and their best competition performances, likewise with their worst. Also, participants who took more warmup vaults performed significantly better on average. Athletes and coaches should consider implementing warmup practices that emphasize familiarizing oneself with their performance environment.

KEYWORDS
Pole Vault; Track and Field; Warmups; Warmup Performance; Competition Performance; Performance Environment; Nested Task; Task Constraints

INTRODUCTION

Compared to practice, competitions often require athletes to perform under new or more challenging circumstances. The competition conditions external to the athlete constitute the performance environment, and can influence multiple skills relevant to athletic performance such as perceiving distance,[1,2] memorizing distance,[3] and perceiving velocity.[4] Familiarity enhances one's ability to perform physical tasks such as navigating,[5] coordinating motor responses,[6] and approaching a predetermined target within that environment.[7] This relationship is also evident when considering the phenomena of the "home field advantage" with home teams winning significantly more competitions across multiple sports and leagues.[8] In soccer, this effect was significantly reduced when home teams changed stadiums, suggesting the advantage came from familiarity with the conditions of the arena.[9] This raises the question whether increasing one's familiarity with their performance environment would positively influence their performance within a competition.

Track and field events such as the long jump and pole vault are unique to other events and sports in that athletes are allowed a "warmup period" before the event starts using the same area and performing similar jumps to what they will during the competition. In the case of the pole vault, the warmup period is usually 30-60 minutes long and athletes will take turns using the runway and landing area to perform multiple vaults as they will eventually in the competition. Panteli *et al.* (2016) explored how the environment and task influenced long jumpers' visual step regulation and found that jumpers exhibited better step regulation and accuracy when the task more closely mimicked a full long jump sequence rather than shortened drills.[7] They also exhibited better performance on approach drills when performing this task on the long jump runway they normally used rather than a flat track. These results indicate that familiarity with the environment and executing the complete task leads to better performance of elements within that task. Completion of a vault in the pole vault occurs when the athlete builds speed on the runway, plants their pole into a box, swings upside down, and finally launches themselves into the air in an attempt to clear a horizontal bar. Every one of these distinct stages is vital to a successful pole vault attempt,[10] making the pole vault an ideal event to investigate the importance of task completion.

The current study investigated the relationship between warmup and competition performance in the pole vault. Besides there being no bar to clear during the warmup period, the warmup and competition environments are held constant, making a measure of the completeness of each vault an ideal indicator for warmup performance. Based on the prior research, a higher level of completeness of warmup vaults will indicate better mastery of the competition environment. The total number of warmup vaults will also be considered as repetition contributes to familiarity.[11,12] We hypothesize that better performance during the warmup period, as measured by vault completeness, will be associated with better performance during the competition. Based on the research pertaining to familiarity and performance cited earlier,[5-7,11,12] we also hypothesize that more warm up vaults will be

associated with better competition performance. The findings of this study may provide insight into the relationship between performance and environment within the pole vault and may help coaches increase the effectiveness of practices during the warmup period.

METHODS

Participants

A total of 16 pole vaulters, 7 male and 9 female, participated in the current study and were selected based on convenience. The number of males in their first, second, third, and fourth year of collegiate competition were 4, 0, 2, and 1, respectively (M= 2.00, SD= 1.20). The number of females were 1, 2, 5, and 1, respectively (M= 2.67, SD= .82). All participants were NCAA Division II collegiate athletes attending a midwestern university and ranged from 18-21 years of age at the start of the study. The average personal records, in meters, for males and females at the start of the study were 4.58m (SD= .33) and 3.62m (SD= .33), respectively. This sample demographic was chosen because competition at this level requires a substantial degree of technical proficiency that allowed standardization of measures that were based on form. Written informed consent was obtained from all of the participants. Approval for this study was obtained from the university's institutional review board (reference number: 20-117-H) and athletic department.

Procedure

Testing occurred at the same indoor track and field arena at seven different pole vault competitions that adhered to NCAA rules and regulations. These took place over the course of 45 days. The participants were observed by one of two trained research assistants during the warmup period only, both who had extensive prior knowledge and experience with the pole vault. A test of interrater reliability was conducted where both observers independently scored the same 16 warmup vaults performed by four different participants (four vaults each) during the first pole vault competition of the study. The scores were then compared via Pearson correlation, $r(14) = .97$, $p < .001$. Observers were instructed to interact with the participants as little as possible. Data was collected on each warmup jump performed by each participant. Data was only collected on warmup jumps from the longest approach used by the participants in warmups in an effort to match the data with expected competition conditions. The data was written on a coded template after each vault.

The data recorded during observations consisted of the completeness score of each warmup vault for each participant that was competing that day (**Table 1**). Completeness was measured on a 0-3 scale based on how far the participant proceeded through the stages of the vault: a score of zero represents a vault in which the participant did not leave the ground, a score of one represents when the participant left the ground but made little or no movement to swing the body upside down, a score of two represents when the participant left the ground and swung upside down so their legs went past parallel with the ground, and a score of three represents when the participant left the ground, swung upside down, and made a concerted effort to mimic clearing a bar as they would in a competition (**Figure 1**). An obstacle such as a practice bungee did not need to be present to achieve a score of three.

Participant	**Warmup Vault Number**						
	1	2	3	4	5	6	7
	Completeness Score on 0-3 Scale						
A	3	2	1	3	2	1	3
B	0	0	1	1	2	NA	NA
C	1	1	2	2	3	3	NA

Table 1. Example of fictional observational data. "NA" is used in instances where the participant took less than that number of warmups.

Apart from the data collected during observations, the participants' performance in the competitions as well as their personal records were obtained from the public domain via "athletic.net."

Participants' competition performance was measured by using a ratio of the height scored in competition (CH) to the participant's personal record (PR) at the time of that competition. This allowed for the comparison of all the participants regardless of individual differences in competency or experience. This calculation resulted in a performance score typically ranging from zero (in the case a participant did not make any height) to one (the participant tied their PR). If a participant vaulted a new PR, a score greater than one would result, and the new PR would be used in the calculation for subsequent competitions. This was done to control for any maturation that occurred over the course of the study.

The main analyses consisted of comparing competition performance with multiple variations of the warmup scores via Pearson correlation. Competition performance was also compared with the total number of warmup vaults taken (TWUV). In addition to

the average scores for all of the warmup vaults for a given competition (S), the average of all of the warmup vaults except the first (SE1) and the average of the last three warmup vaults (SL3) were included.

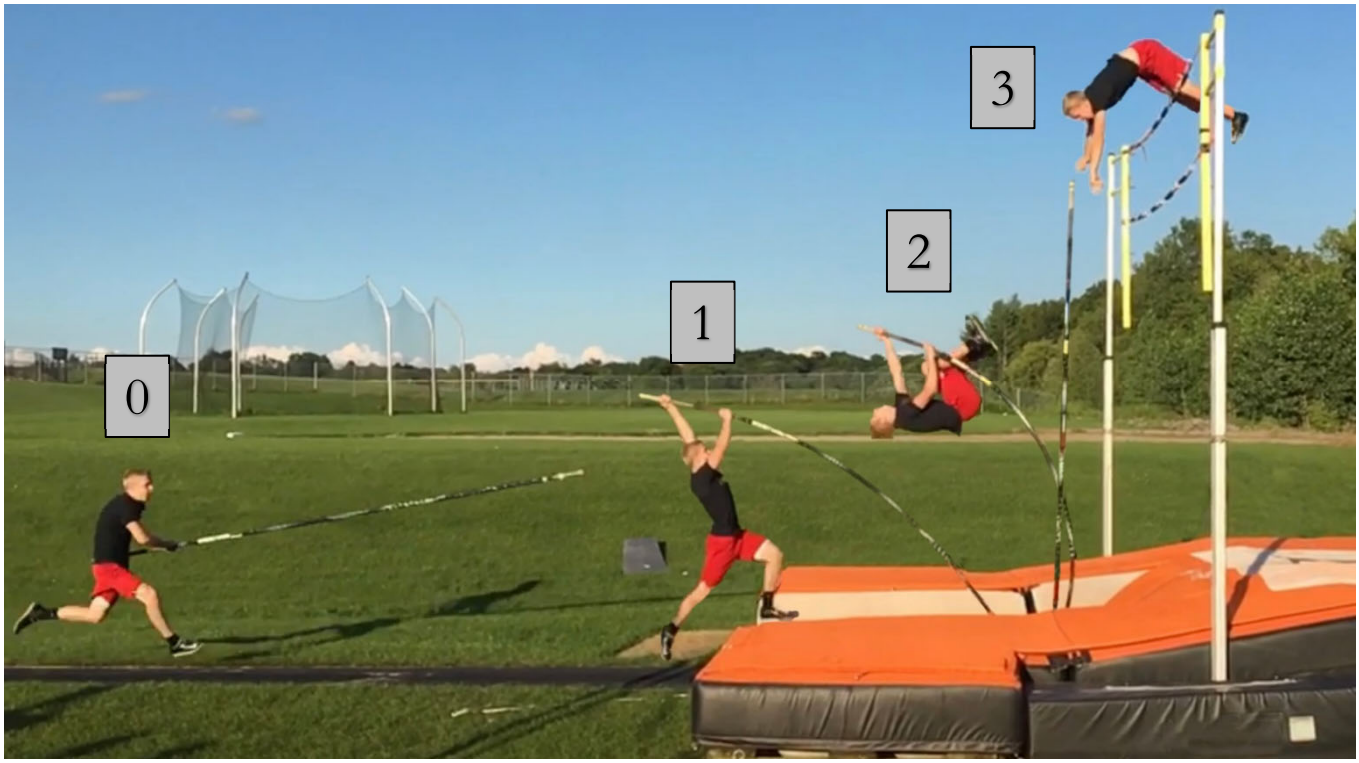

Figure 1. Pole vault stages and corresponding vault completeness scores. (Original photo)

The "SE1" measure was used because a vast majority of all first warmup jumps scored a one, suggesting participants were intending to score this as a form of a warmup drill. The "SL3" measure was used to test if the last three warmup vaults had more influence than the preceding ones. One point to note is that these measures included instances in which a participant did not make any height in a competition, commonly known as a "no height" (NH). Functionally, these scores represent the same as vaulting "zero" in competition. In an attempt to prevent these scores of zero skewing the correlative data, performance and scores were averaged for each participant then compared across participants.

An alternate analysis was then performed to address NH data in a different fashion that represented them as individual observation results without skewing the data set. This was done by ordering all of the performances for each athlete from highest to lowest (PerfRank) and comparing that order to the corresponding order of warmup scores (SRank), warmup scores except the first (SE1Rank), last three warmup scores (SL3Rank), and the total number of warmup vaults (TWUVRank) via Spearman's rank correlation.

RESULTS

Across 16 participants over the course of seven competitions, a total of 60 observations took place, with 15 resulting in a NH. The Pearson correlation between performance and TWUV (**Figure 2)** was significant, while no significant relationships were found between performance and Score, ScoreE1, or ScoreL3 (**Table 2**).

When comparing the data rank wise, significant Spearman's correlations were found between PerfRank and SRank, as well as PerfRank and SE1Rank, while no significant relationships were found between PerfRank and SL3Rank or TWUVRank (**Table 2**).

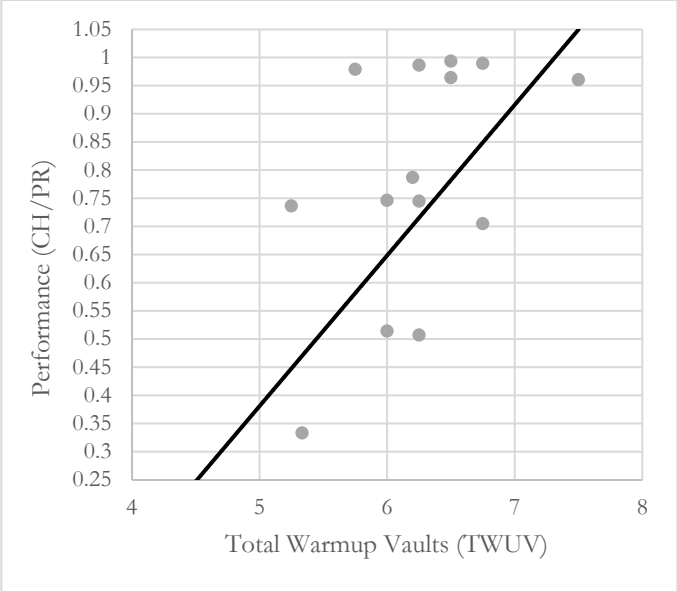

Figure 2. Scatterplot and trendline showing Pearson correlation between Total Warmup Vaults and Performance.

Variable	Performance		Variable	Performance Rank	
	r^a	p		ϱ^b	p
S	.35	.189	**SRank**	.31	.014*
SE1	.36	.172	**SE1Rank**	.32	.012*
SL3	.43	.099	**SL3Rank**	.23	.078
TWUV	.56	.029*	**TWUVRank**	.19	.148

Table 2. Correlative results. r – Pearson correlation coefficient, p – level of significance, ϱ – Spearman's correlation coefficient, S – average score of all warmup vaults, SE1 – average score of all warmup vaults except the first, SL3 – average score of the last three warmup vaults only, TWUV – total warmup vaults, Rank – rank order highest to lowest, * - Significant result ($p<.05$), [a]Degrees of freedom = 14, [b]Degrees of freedom = 58.

DISCUSSION

The results of this study support the hypothesis that better performance during warmups, as measured by vault completeness, is associated with better performance in competition. They also support the hypothesis that taking more total warmup vaults is associated with better competition performance.

Participants' best scores during the warmup period significantly correlated with their best competition performances. This was true when comparing the average of all the warmup jumps in a given competition as well as the average of all except the first warmup jump. When considering the number of warmup vaults taken before competition, participants who took more warmup jumps, on average, performed better in competitions over the course of the study. Had competition performance been a measure of the height vaulted, this result could easily have been attributed to confounding differences between participants. However, because performance was measured as a ratio comparing participants' performance to their own PR, this explanation is less likely.

One important feature to consider is the role NHs take within the interpretation of the results. NHs present a unique problem in that they are a qualitative representation of performance while all other marks are quantitative. This makes them challenging to draw meaningful comparisons between the associated data. To incorporate NHs as vaulting a height of "zero" would skew the data, while removing them from the data set altogether would lead to the consideration of only a "higher performance" subset of what actually occurs in competition. Either of these could explain why the Pearson correlations between performance and warmup scores failed to reach significance while the Spearman rank order correlations did. Including NHs in the average performance of each vaulter does create a useful means of comparison, but it also eliminates their significance as distinct observations which may lead to their influence being underrepresented. In light of this, there was still a significant Pearson

correlation between performance and TWUV suggesting the effect was robust enough be present despite the data being averaged. In a similar manner, the Spearman's correlations between PerfRank and TWUVRank as well as SL3Rank were insignificant. This is likely due to the nature of the variables themselves. With TWUV ranging from 5-8, ranking led to a large number of ties. The same was true with SL3, where ranking the average of only three vaults led to a large number of ties.

It is worth noting that, while some of the relationships did reach significance, their correlation coefficients were not particularly high. The variables were compared using linear correlation, but it is entirely possible these relationships are more complex and require different analyses. Perhaps with longer observation periods constituting multiple pole vault seasons and a larger sample, the relationships will be better characterized for consideration of non-linear comparisons of the variables.

Low warmup vault scores suggest that a participant was unable to complete vaults due to inadequate performance of the preceding stages of the attempt. This is likely due to unfamiliarity with aspects of the competition environment relevant to proper execution of the earlier stages of the vault *e.g.* reaching a desired take-off target with unfamiliar cues.[7,13] It may also be due to conditions external to the physical competition but still within the environment itself such as crowd size, temperature, time of year, stakes of the competition, skill of the competitors, etc. When considering the high level of experience participants have with the event, an environmental explanation becomes even more likely. This would suggest that those who had high warmup vault scores not only had higher environmental familiarity, but that this translated to better performance in competition, as the results support. In this context, it is worth noting that all of the data in the present study came from the same facility, an indoor fieldhouse which is the participants' most common place of competition (i.e., their "home" location). A valuable direction of future research would be to repeat this study at different competition venues to see if the effects are enhanced in alternate environments.

The main limitations of the study come from the challenge presented by analyzing NH data. To limit skewing the data, performances were averaged within participants at the expense of the representativeness of each individual observation. To combat this, performance was analyzed in terms of rank. This allowed the representation of NH data, but at the expense of analyzing relationships based on the magnitude of difference. A future direction of research could be collecting more data and using it to create a model by which NHs can represent a nonzero performance that can be meaningfully compared to others. Collecting more data may also help address the correlation coefficients of the significant findings and help to further characterize their relationship. To avoid adding any potential stress to the athletes, because data was collected in NCAA sanctioned competition, active involvement of participants in the research by way of surveys/questionnaires was not requested. Because of this, other factors such as weight, length, and flex of poles being used, documenting when each athlete switched poles, reflecting on warmups/competition performance, etc., were not considered. Future research observing athletes in unofficial competitions or practice would allow this limitation to be addressed.

CONCLUSION

To more effectively utilize warmup periods, coaches and athletes should consider adopting practices that focus on familiarizing oneself with their competition environment or simulating the competition environment during practice. Some examples of this may include: using the same runway and landing area that will be used in competition, recreating potential competition scenarios, taking more vaults from a full approach, using a competition bar in practice instead of a bungee, setting the standards to what an athlete normally uses in competition, and giving coaching instructions as one would in a competition. While these results do provide valuable insight into pole vault performance, it should be acknowledged that participants in the study constituted only a sample and attended the same university, experiencing similar training and coaching. When considering warmups, participants' best vault completeness scores coincided with their best competition performance scores. Also, participants who took more warmup jumps on average performed better in competition.

ACKNOWLEDGEMENTS

The author thanks Dr. Robert Deaner for his guidance and continued interest in the development of the author's research skills and knowledge throughout this project. The author also thanks Ellianne Martin, whose involvement was integral to the completion of this work.

REFERENCES

1. Lappin, J. S., Shelton, A. L., & Rieser, J. J. (2006) Environmental context influences visually perceived distance. *Perception & Psychophysics*, 68, 571–581. *https://doi.org/10.3758/BF03208759*
2. Witt, J. K., Stefanucci, J. K., Riener, C. R., & Proffitt, D. R. (2007) Seeing beyond the Target: Environmental context affects distance perception. *Perception*, 36(12), 1752–1768. *https://doi.org/10.1068/p5617*

3. Morone, G., & Paolucci, S. (2012) Walking there: Environmental influence on walking-distance estimation. *Behavioural Brain Research*, 226(1), 124.
4. Epstein, W. (1978) Two factors in the perception of velocity at a distance. *Perception & Psychophysics*, 24(2), 105–114. *https://doi.org/10.3758/BF03199536*
5. Muffato, V., & Meneghetti, C. (2020) Knowledge of familiar environments: Assessing modalities and individual visuo-spatial factors. *Journal of Environmental Psychology*, 67, 101387. *https://doi.org/10.1016/j.jenvp.2020.101387*
6. Pinder, R. A., Davids, K., Renshaw, I., & Araújo, D. (2011) Manipulating informational constraints shapes movement reorganization in interceptive actions. *Attention, Perception and Psychophysics*, 73(4), 1242–1254.
7. Panteli, F., Smirniotou, A., & Theodorou, A. (2016) Performance environment and nested task constraints influence long jump approach run: A preliminary study. *Journal of Sports Sciences*, 34(12), 1116–1123. *https://doi.org/10.1080/02640414.2015.1092567*
8. Jamieson, J. P. (2010) The Home Field Advantage in Athletics: A Meta-Analysis. Journal of *Applied Social Psychology*, 40(7), 1819–1848. *https://doi.org/10.1111/j.1559-1816.2010.00641.x*
9. Pollard, R. (2002) Evidence of a reduced home advantage when a team moves to a new stadium. *Journal of Sports Sciences*, 20(12), 969–973. *https://doi.org/10.1080/026404102321011724*
10. Iosa, M., Fusco, A., Gudelj, I., Zagorac, N., & Babić, V. (2013) Influence of Kinematic Parameters on Pole Vault Results in Top Juniors. *Collegium Antropologicum*, 37 Suppl 2, 25–30.
11. Johnson, H. W. (1961) Skill = speed × accuracy × form × adaptability. *Percept. Mot. Skills* 13, 163–170.
12. Lee, T., Swanson, L., & Hall, A. (1991) What Is Repeated in a Repetition? Effects of Practice Conditions on Motor Skill Acquisition, *Physical Therapy*, 71 (2) 150–156, *https://doi.org/10.1093/ptj/71.2.150*
13. Needham, L., Exell, T. A., Bezodis, I. N., & Irwin, G. (2018) Patterns of locomotor regulation during the pole vault approach phase. *Journal of Sports Sciences*, 36(15), 1742–1748. *https://doi.org/10.1080/02640414.2017.1412236*

ABOUT THE AUTHOR

Alex Peskin graduated from Grand Valley State University in April of 2020, dual majoring in Biomedical Science and Behavioral Neuroscience. He currently works as an oncology research coordinator and plans to pursue a medical degree.

PRESS SUMMARY

The pole vault is a track and field even where athletes use a long pole to launch themselves high in the air over a bar. In a pole vault competition, athletes have the opportunity to warm up in the same area and perform the same motions they will in the competition. The aim of this research was to determine if there was a link between an athlete's performance in warmups and in competition. After observing collegiate pole vaulters and tracking their performances, a link was found. Athletes who performed better in warmups also performed better in the competition. Also, athletes who took more practice runs performed better. These findings indicate that certain aspects of the warmup process could be important to an athlete's success in a competition. If athletes and coaches were to focus on key factors such as simulating the competition environment in practice and focusing on completing warmup vaults, they may have better performances.

A Narrative Literature Review of the Psychological Hindrances Affecting Return to Sport After Injuries

Ashley Sweeney[a], Stephanie M. Swanberg*[b], & Suzan Kamel-ElSayed*[c]

[a]*Honors College, Oakland University, Rochester, MI*
[b]*Michigan School of Psychology, Farmington Hills, MI*
[c]*Department of Foundational Medical Studies, Oakland University, Rochester, MI*

https://doi.org/10.33697/ajur.2021.045

Student: asweeney@oakland.edu, ashsweeney25@gmail.com*
Mentors: sswanberg@msp.edu, & elsayed@oakland.edu

ABSTRACT
After different sports injuries, athletes may experience various psychological emotions in response to such injuries, which could lead an athlete to feel stressed. These emotions include anger, fear, frustration, anxiety, and depression which may lead to lack of confidence in returning to their sport and/or fear of sustaining a new injury. This narrative review aims to determine the possible psychological hindrances present when an athlete is planning on returning to sport after injury to an anterior cruciate ligament (ACL) or after sustaining a concussion. The synthesized information for this review has been collected from researching the databases PubMed, SportDiscus, PsycInfo, and Google Scholar using search terms including "return to sport", "ACL injury", "concussion", and "psychology". Journal articles needed to be in English and published in the years 2009-2019; books and unpublished abstracts were excluded. A total of 42 studies were included and analyzed using deductive coding to organize and synthesize relevant articles into themes. The review summarizes the shared common and the different psychological hindrances that may be found in athletes after an ACL injury or concussion. Shared psychological characteristics for returning to sport following either an ACL injury or concussion included fear, self-esteem, control, anxiety, stress, recovery, and social support. Discovering the common and unique psychological barriers which may affect the injured athletes from returning to sport can help educate athletes' families, coaches, and healthcare professionals, as well as promote discussions for the future to help athletes feel more secure in their return to their respective sport.

KEYWORDS
ACL Injuries; Concussions; Sport Injuries; Athletes; Narrative Literature Review; Psychological Hindrances; Psychological Characteristics; Return to Sport; Psychology

INTRODUCTION

Sports have become an ever-dominant presence in our culture with many people choosing to play sports at the high school, collegiate, professional, or recreational level. Specifically, over 8 million high school students and nearly 500,000 college students participate in athletics as of the 2018-2019 year while 19.5% of US adults participated in some form of daily exercise.[1-3] Being involved in sports comes with the increased chance of injury, especially anterior cruciate ligament (ACL) injuries and sports related concussions.

ACL injuries are one of the most common sports related injuries in the United States with roughly 130,000 ACL reconstructions performed by doctors annually.[4] Many athletes that sustain ACL injuries often undergo some type of rehabilitation, most commonly physical therapy, but sometimes this rehabilitation may come with a heightened risk of return to sport (RTS). Research suggests that 50% of patients that undergo ACL reconstruction may not be able to reach their previous level of sport, prior to injury.[5,6] Although adequate rehabilitation and successful reconstruction are attainable in most athletes, RTS may not always be achievable since there are many other factors affecting RTS. They could be related to physical factors, (e.g., impairments an athlete may feel after sustaining their injury), contextual factors (e.g., social, and environmental pressures), and their own psychological hindrances they may be experiencing.[7]

Concussions pose another threat to the athlete, as this injury has many hidden symptoms. Since there are no outward physical impairments for concussions, it can be very difficult to predict how long it will take an athlete to recover.[8] From 2010-2016, there were over 280,000 hospital visits per year related to traumatic brain injuries, with about 45% of those occurring from contact

sports. [9] While this is deemed as the "invisible" injury, the jolting of the brain when a concussion occurs can cause symptoms that may include headaches, anxiety, irritability, and memory impairments.[10]

Greater public concern around concussions has emerged as more health professionals and researchers gain a better understanding of how concussions affect the brain.[11] Not only have concussions become a central focus for healthcare professionals, but also become aware of how concussions also impact parents and families of athletes who sustained these injuries. This awareness has prompted all 50 US states to pass laws to help safeguard athletes and support further research.[12]

This literature review provides an in-depth analysis of the psychological hindrances that affect RTS after sustaining an ACL injury or concussion. The focus of this review is on these two athletic injuries, as both injuries can have lengthy recovery times, complications, and similar impacts on the psychological health of athletes. Other sports injuries were excluded by the authors to provide a narrow focus and allow for more concise data collection and synthesis.

METHODS

This literature review followed the matrix method of conducting literature reviews described by Goldman and Schmalz (2005). [13] Inclusion criteria included English language journal articles published from (2009-2019) which summarized findings related to the psychological factors that may hinder athletes' return to sport after an ACL injury or concussion. Any age of athlete was also included. Exclusion criteria included injuries other than ACL or concussion. The databases of PubMed, SportDiscus, PsycInfo, and Google Scholar were searched using the search terms listed in the table below **(Table 1)**.

Database	Search Strategy
PubMed	(((Anterior Cruciate Ligament OR ACL) AND (injury OR injuries OR tear OR tears OR reconstruction)) OR "Anterior Cruciate Ligament Injuries"[Mesh] OR concussion* OR "Brain Concussion"[Mesh])) AND (fear* OR anxiety OR anxious OR "Fear"[Mesh] OR "Anxiety"[Mesh]) AND ("return to sport" OR "return to play" OR "Return to Sport"[Mesh])
PsycINFO	(((Anterior Cruciate Ligament OR ACL) AND (injury OR injuries OR tear OR tears OR reconstruction)) OR concussion*) AND (fear* OR anxiety OR anxious) AND ("return to sport" OR "return to play")
SportDISCUS	((Anterior Cruciate Ligament OR ACL) AND (injury OR injuries OR tear OR tears OR reconstruction) OR concussion*) AND (fear* OR anxiety OR anxious) AND ("return to sport" OR "return to play")
Google Scholar	(ACL injury* OR concussion*) AND (fear or anxiety) AND ("return to sport" OR "return to play")

Table 1. Search Strategies.

Following the searches, the primary author reviewed and selected studies to include in the literature review based on the criteria. Each article was separated into either the ACL and concussion category, then divided into sub-categories or themes using the matrix method to organize and summarize the literature. The sub-categories were determined prior to data collection, followed by organizing the data through deductive coding. Once the categories were determined, each article was read a second time to determine the most relevant themes. For ACL injuries, the articles were grouped in the following categories: self-reported fear and fear of re-injury, and psychological factors hindering return to sport. For concussions, the articles were separated into the following categories: recovery patterns after sustaining a concussion; anxiety and social support post-concussion; depressive symptoms in concussed athletes and return to play after sustaining a concussion. No assessment of the methodological quality of the articles was conducted.

RESULTS

A total of 43 studies were included in this literature review. The following section summarizes the findings of these studies on ACL injuries and concussions, respectively.

ACL Injuries

Self-Reported Fear and Fear of Re-injury

The topic of fear and fear of re-injury was a common theme that emerged in this review. To determine an athlete's self-reported fear at the time of return to sport (RTS), Paterno *et al.* (2018) created a study in which they compared self-reported fear as well as the incidence of a second ACL injury within 24 months of that athlete sustaining their first injury[14]. Categories were created for individuals that had undergone ACL reconstruction (ACLR) ranging in ages from 10-25 years of age. In the Paterno *et al.* study, fear was measured using the shortened version of the Tampa Scale of Kinesiophobia (TSK) [15] with the participants separated into two groups. In the TSK scale, those who scored a 17 or greater on the TSK-11 are categorized in the high fear group and those that scored 16 and lower were in the low fear group.[15] Results of the study showed that those who reported a greater fear on the TSK-11 were up to four times more likely to show lower levels of activity once they had undergone ACLR. Comparatively, those who had returned to their normal states of performance prior to injury often reported higher levels of fear when demonstrating

pivoting and cutting motions. This higher self-reported fear also led to those athletes having a higher risk of suffering a second ACL injury within the 24-month period in which the participants stayed in contact throughout the study.[14]

Tripp *et al.* (2011) conducted a similar study about confidence in RTS for athletes one year after having undergone ACLR.[16] The TSK-11 scale was used to measure the psychological aspects of the participant returning to sport and the Shortened Profile of Mood States (S-POMS) was used to measure mood states of those who had undergone ACLR.[15,17] The S-POMS consists of 37 topics for six different emotion categories which include: tension/anxiety, depression/dejection, anger/hostility, fatigue/inertia, vigor/activity and concussion/bewilderment.[17] The Pain Catastrophizing Scale (a 13-item scale) was also used to ask participants questions about their thoughts and feelings in regards to the pain after ACLR.[18] An example of a question that would be asked on the scale is "when I'm in pain, I worry all the time about whether the pain will end." [18]

Based on the Tripp *et al.* study, confidence was assessed to determine how ready an athlete was to RTS. [16] To measure confidence, the Sport Self-Confidence Inventory, which consisted of 13 different topic areas, was used to assess the confidence participants had about participation in sports.[19] Those that scored a 1 were said to have low confidence and those that scored a 9 had high confidence.[19] The results of this study showed that those who had a high fear of re-injury and negative mood states after ACLR were more likely to have low confidence in RTS. This correlation was evident in those that reported a higher fear of sustaining a second ACL injury and were more likely to see lower levels of RTS in their respective sport(s).[16]

Psychological Factors Affecting Return to Sport

Nwachukwu *et al.* (2019) conducted a study in which they researched the different factors that affected RTS.[20] They found factors that affected RTS both at the pre-operative and post-operative level for athletes, which seemed to vary. At the pre-operative level, they found that motivation of RTS, importance of RTS, and the possibility of RTS played a large role in the psychological readiness for an athlete to return to their sport. At the post-operative level, they found other variables that played a factor in RTS. These variables were the motivation they felt during rehabilitation such as physical therapy, self-esteem throughout the recovery process, their locus of control (in terms of their health) and also varying states of mood.[20]

Similarly, a study conducted by Ardern *et al.* (2013) found that positive or negative mental states may influence athletes' decisions in RTS. [21] If an athlete has a positive psychological response to surgery and post-operative rehabilitation, using crutches or physical therapy, they are more likely to RTS after a 12-month period than those who have a negative psychological response to RTS.[21] While a patient is undergoing rehabilitation, it is essential to facilitate a positive psychological response to ensure a quicker RTS.

There are instances in which an athlete may sustain a second ACL injury and experience similar or heightened psychological reactions or emotions from the first injury. Webster *et al.* (2018) conducted a study in which they focused their research on psychological readiness when an athlete has sustained a second ACL injury after undergoing ACLR for their first injury.[22] Psychological readiness was determined by the Anterior Cruciate Ligament Return to Sport After Injury Scale (ACL-RSI).[23] This scale measures an athlete's emotion, confidence, and risk appraisal. For independent measures, there were five variables that were measured: surgical timing (i.e., the timing in between when an athlete sustained their injury to the time they underwent surgery), pre-injury sports participation (i.e., how active the athlete was in their sport before the injury occurred), knee laxity (i.e., how much an athlete is able to bend their knee), limb symmetry index (i.e., how similar each athlete's knees are) and subjective knee symptoms and functions, reported by the patient.[22]

The results of Webster *et al.* (2018) suggested that young males athletes had a decreased time between injury and surgery, had a higher level of pre-injury sports participation, a higher limb symmetry, higher knee scores, and reported a higher psychological readiness for RTS after sustaining a second ACL injury, when compared to young female athletes, suggesting sex differences.[22] These higher subjective knee scores combined with a confidence to ensure proper rehabilitative outcomes is also correlated with a higher psychological readiness to RTS. These sex differences have also been demonstrated in another study by Ardern and colleagues about men that have a higher psychological readiness to RTS.[6] They also suggested that male athletes are also more likely to return to their previous level of competition before their injury occurred compared to female athletes.[6]

While there are many studies that have been previously published that focus on why an athlete may be experiencing the emotions post-injury, there may be another explanation about this "fear" that athletes experience when they RTS. Walker *et al.* (2010) conducted a study in which they focused on the topic of "re-injury anxiety".[24] This concept centers around why fear may not be the proper term to use for what an athlete is experiencing when they RTS. Walker and colleagues stated that fear is more of a biological response, but anxiety is more of the anticipatory feeling that an athlete may experience when they are contemplating the timeframe in which they should RTS.[24] This could also correlate with psychological readiness; if an athlete reports more of the "re-injury anxiety", then they may exhibit lower levels of psychological readiness to RTS.

Further research has demonstrated there are more psychological factors affecting readiness to RTS, including self-efficacy, stress, and social and contextual factors. A study conducted by Everhart *et al.* (2015) focused on three psychological domains that affect RTS: fear avoidance, self-efficacy, and stress and health.[25] To measure fear, the Fear Avoidance Model, was used in which they found those who had a negative outlook response to the injury also had a negative emotional response to their ACL injury.[26] To measure self-efficacy, the Theory of Self Efficacy was used.[27] Using this measurement, it was found that having a patient set goals and portray positive self-talk performed better in their rehabilitative measures and thus had a higher level of RTS.[25] Stress and health were measured by the Stress, Health and Buffering Hypothesis.[28] It was found that individuals with higher stress combined with social support had better exercise completion both on their own and in physical therapy. However, to achieve these positive results, athletes must be willing to push past the fear of sustaining another injury to their ACL and be willing to return to their sport at the same level of functioning pre-injury.[21]

A second article by Ardern *et al.* (2015) focuses on the physical, social, and contextual factors that play into psychological readiness for an athlete to RTS.[29] Specifically, sociodemographic factors such as age, sex, ethnicity, and socioeconomic status can play a larger factor in RTS.[29] Ultimately, this could hinder an athlete's rehabilitative outcome and may lead to a decreased level of psychological readiness. Finally, another limitation may stem from the overall performance of an athlete's knee, focusing specifically on motor control and balance.

Concussions

Concussions are injuries that occur inside the brain hence researchers cannot see them properly to fully diagnose and understand them. [30] There is a plethora of symptoms that can be seen and used to gauge a greater understanding of what occurs inside of the brain after it has been jolted by an outside force. Concussions can have physical symptoms, which can stem from headaches to feelings of dizziness, cognitive problems (e.g., difficulty with memory, the inability to concentrate, etc.), trouble sleeping, and emotional disruptions (e.g., irritability and anxiety). [30] These symptoms will be analyzed in further detail in the following sections.

Recovery Patterns After Sustaining a Concussion

Recovery patterns after an athlete sustains a concussion can often be difficult since there are a variety of symptoms an athlete can possess. These symptoms are also based on the athlete's self-report, which may not be reliable and include individual differences since self-reports tend to be biased as people like to present themselves in a more favorable demeanor. A study conducted by Teel *et al.* (2017) focused on some of the symptoms that may occur after sustaining a concussion.[31] They determined that an athlete that had loss of consciousness (LOC) may experience amnesia, which may occur in two types: post-traumatic amnesia and retrograde amnesia. Athletes may experience post-traumatic amnesia in which they would not be able to remember anything after sustaining the concussion. Other times, they may experience retrograde amnesia, where athletes would not be able to remember anything prior to the concussion occurring.[31] This information was used in augmentation to three additional measures used to assess concussion symptoms, psychological assessments, and balance which are detailed below.

The Graded Symptom Checklist (GSC) was used to determine symptoms, ranging on a scale from 0-6 with 0 meaning no symptoms and 6 being the most severe.[32] This measure recorded four different types of symptoms: somatic (early and evolving onset), cognitive, and neurobehavioral symptoms. Early onset somatic symptoms could be dizziness, headaches, nausea, and vomiting. Evolving onset somatic symptoms could include drowsiness and fatigue. Cognitive symptoms could be having a hard time concentrating and remembering, and neurobehavioral symptoms could range from difficulty sleeping to sadness. [31]

The Standardized Assessment of Concussion (SAC) was used to measure concentration, orientation to the room and objects, immediate and delayed memory, sensation, strength, and coordination.[33] This was measured in a range of 0-30, with a lower score indicating that the athlete possessed worse symptoms. The Balance Error Scoring System (BESS) was used to measure balance on each individual, as well as both legs.[34] This was measured with a range of scores from 0-60, with the higher the score denoting worse balance.

These assessments showed that those who had previously suffered from LOC and suffered from 0-30 minutes of amnesia were more likely to have worse symptoms in any category. Those who had over thirty minutes of amnesia had worse balance while athletes who sustained more than two concussions in their athletic history also exhibited worse balance overall.[31]

Anxiety and Social Support After Sustaining a Concussion

Psychological traits and symptoms can impose as much of a hindrance on an athlete after a concussion as if they were to sustain an injury that would require surgery. A study conducted by Covassin *et al.* (2014) focused on anxiety and the amount of social support an athlete receives from anyone they may interact with after they have sustained their concussion.[35] This was also compared with the social support that an athlete may experience if they have undergone an orthopedic surgery, such as an ACLR.

The State-Trait Anxiety Inventory (STAI) was used to determine how much anxiety an athlete experienced.[36] For this, there were two types: state anxiety and trait anxiety. State anxiety is measured by how much an athlete experiences anxiety in the moment about different scenarios they may be presented with during rehabilitation and RTS. Trait anxiety measures how much anxiety an athlete feels in general about different scenarios.[36] The 6 Item Social Support Questionnaire (ISSQ) was used to assess the amount of social support an athlete felt, which was broken into two parts.[37] Athletes were first asked to provide information about the social support they received and were then asked to rate the overall satisfaction from this social support.

The results were quite similar when compared with athletes that had an orthopedic surgery. Athletes who had sustained a concussion received social support from family (89%), friends (78%), teammates (65%), physical therapists (48%), coaches (47%), and doctors (35%). Those who had undergone an orthopedic surgery were also relatively similar, with their highest amount of social support also coming from family and friends. Having positive social support after a concussion also allows for the athlete to reduce their stress. A positive perception of the social support they received also influenced the reduced levels of anxiety.[35] It was also found that athletes after a concussion were more likely to display trait anxiety.

Depressive Symptoms in Concussed Athletes

Along with stress and anxiety an athlete may incur after sustaining a concussion, there is also the possibility for depressive symptoms to become prevalent. A study conducted by Roiger *et al.* (2015) showed the results of depression in patients that had received a concussion.[38] They used the Center for Epidemiological Studies Depression Scale (CES-D) to assess depressive states, one week, one month and three months post-concussion.[39] They found that after the timeframe of one week, athletes showed the most depressive symptoms. These symptoms decreased over time between the one week and one-month time frame. However, none of these levels posed these athletes to be at risk for a clinical diagnosis for depression.[38]

A study conducted by Guo *et al.* (2018) also found similar findings as Roiger and colleagues.[40] This research compared depressive symptoms to those that had undergone orthopedic injuries and found that athletes that had sustained a concussion also showed higher rates of depressive symptoms. They also used the same measure, the CES-D, to assess the depressive symptoms in concussed athletes.[40]

Post-Concussion Return to Play Expectations

Similar to ACL injuries, there is also a certain timetable deemed both safe and suitable for an athlete to return to their respective sport. Since concussions are a closed head injury, it makes it more difficult to determine when an athlete should be cleared to play. A study conducted by Kelly and Erdal (2016) aimed to identify some characteristics, illustrated below, that may provide more insight to the timeframe in which athletes can RTS.[41] The State-Trait Anxiety Inventory (STAI) was used to determine anxiety post-concussion.[36] The Neurobehavioral Symptom Inventory (NSI) was used to measure traumatic brain injury (TBI) symptoms, as well as the Illness Perceptions Questionnaire-Revised (IPQ-R) to assess how the concussion affected the athlete's life.[42-43] The results of this study illustrated that athletes with a history of concussions were more likely to exhibit more post-concussion symptoms.[41] If an athlete also identified as being more athletic prior to their concussion, they also were seen to exhibit less days in between when they sustained the injury and when they decided to RTS.[41] It was also found that amnesia, the type of concussion symptom, and LOC played a significant role in an athlete's decision to RTS.[31]

Concussions being a closed head injury are subject to many self-reported symptoms or instances where an athlete may say they are ready to return to their sport when they inherently may not be. Lower rates of fear of re-injury have been shown to be present in athletes, post-concussion due in part to self-report.[40] Anxiety has also been shown to decrease over time because the injury may not be taken as seriously as an orthopedic injury, such as that of an ACL injury.[40]

Summary of Results

This review discovered both similarities and unique psychological hindrances for athletes suffering from ACL injuries or concussions when returning to sport. **Table 2** below illustrates the psychological hindrances that are present in ACLs and concussions based on the information that has been discussed in the results section. While each respective sport contains hindrances specific to the injury, there is still some overlap. This is illustrated by the third column of the table which illustrates the shared psychological hindrances present in both injuries.

Injury	Psychological Hindrances	Shared Characteristics
ACLs	• Motivation to return to sport • Importance of return to sport • Possibility of return to sport • Mood • Emotion • Self-efficacy • Physical impairments • Quality of life • Severity of injury	• Self-reported fear • Fear of re-injury • Self-esteem • Locus of control • Anxiety • Stress • Recovery patterns • Social support
Concussions	• Amnesia • Concentration • Coordination • Headaches • Dizziness • Depression • Pre-injury History • Cognitive Risks (CTE and Dementia)	

Table 2. The shared common and different psychological hindrances that may be found in athletes after an ACL injury or concussion.

DISCUSSION

Based on the findings from this literature review, there are many psychological hindrances that can affect RTS after an athlete has sustained an ACL injury or a concussion. Though every athlete's experience is unique, level of fear and anxiety were shown to be the two strongest predictors in both injuries when deciding when to RTS.

According to Ardern *et al.* (2015), the severity of the injury which is perceived by the individual may affect their psychological readiness, especially in rehabilitative measures, such as physical therapy.[29] If a patient feels that their injury is severe, they may not want to return on the specified timetable their doctor has advised them to follow or they may personally hinder themselves by not taking rehabilitation as seriously as they should, delaying their RTS.

Motivation also plays an active role in how prepared an athlete feels to RTS. This is exhibited throughout the recovery process from rehabilitation to RTS.[20] The motivation to return can also be hindered by the severity of the injury mentioned above. If an athlete feels their injury is too severe, this may impede the return process and in turn decrease their motivation to RTS.

Though not the focus of this review, return to sport is a complex process that can be affected by not just psychological, but also physical and social impairments. The findings of this review support those reported previously by Ardern *et al.* (2015).[29] Physical impairments may also accompany an athlete's mental impairments. Muscle strength, pain, knee stability, knee swelling, and knee movement can all play a big role in how much an athlete feels they are able to RTS.[29] If an athlete has physical hindrances that inhibit them in their RTS, this will inherently influence their mental state and further inhibit them on their RTS. **Figure 1** summarizes commonalties between this review and Ardern *et al.* (2015). Psychological factors that affect RTS following ACL injuries including physical impairments, such as muscle strength, pain, stability of the knee, swelling of the knee, and the overall movement of the knee. Social and contextual factors may differ between athletes, some being too optimistic or too pessimistic about the recovery outcomes. All these factors are intertwined with one another, as well as combined with others. Certain variables such as the cause and severity of the injury could play a large role in an athlete's psychological readiness to RTS.

Figure 1. Biopsychosocial model of return to sport after an ACL injury. Adapted from Ardern, C. L., Kvist, J., & Webster, K. E., Biopsychosocial model (2015).[29]

Throughout this literature review, the research has illustrated that there are many psychological hindrances that affect RTS after an athlete has sustained an ACL injury. Concussions also pose their own set of psychological hindrances when an athlete is considering returning to their sport.

Depression plays a substantial role in when an athlete can RTS. Depressive symptoms were shown to be higher in athletes who sustained a concussion compared to those that underwent orthopedic injuries.[40] While these athletes show depressive symptoms, elevated from normal depressive levels, these levels are not a cause for concern for the athlete. Levels are not high enough for a clinical diagnosis and will return to their normal levels once the athlete returns to their sport.[38] While depression can be a hindrance in RTS, it can be overcome with time.

Pre-injury history can also influence RTS after sustaining a concussion. If athletes had a history of prior concussions, they showed a higher likelihood of having post-concussion symptoms.[41] These symptoms could include depressive symptoms as mentioned above. If an athlete has prior concussions, they also may experience anxiety that would hinder them even further in the return process.

Those that sustain an ACL injury were seen to have higher levels of fear that were associated with being four times more likely to show lower levels of activity.[14] This is illustrated after the athlete has undergone ACLR and is in the process of RTS. Concussed athletes also had self-reported fear. If they self-reported symptoms such as that of LOC or amnesia, athletes were shown to have a lower time frame in which they returned to sport.[31]

Fear of re-injury is very common in athletes who have sustained an ACL injury. However, if an athlete was able to set goals and possessed positive self-talk, they were shown to have higher levels of RTS.[25] In a similar manner, athletes that sustained a concussion and had lower self-reported fear, had a lower fear of re-injury.[40] This could also factor into a concussion being an invisible injury since there may not be any physical symptoms seen in concussed athletes.

Anxiety and stress are both detriments to the return process after sustaining an ACL injury or concussion. For ACL injuries, even if athletes had a higher stress level, if they possessed a high level of social support, they possessed better physical therapy outcomes. These combined factors allowed for an overall lower stress level when deciding to RTS. However, an athlete would have to push past the fear of a second ACL injury to return to their previous level, meaning they would need to surpass their levels of stress.[25]

For concussions, high trait anxiety was shown to decrease and aid in an athlete's RTS.[40] Similarly, to ACL injuries athletes that sustained concussions but had a higher level of social support were able to reduce their stress levels. Athletes that exhibited a positive perception of social support displayed an overall reduction level in anxiety.[35]

CONCLUSION

ACL injuries and concussions are some of the most prominent injuries in sports at any level. This literature review summarizes the psychological hindrances that affect RTS after sustaining an ACL injury or concussion and provides future directions that could be beneficial to the athletic community including coaches, parents, educational institutions, and the athletes themselves. In addition, this review could bring more awareness of the similar and unique psychological factors presented by ACL injuries and concussions inhibiting an athlete's return to sport. The summary list of potential psychological factors hindering athletes from returning to sport could allow the athletic community to be better equipped to provide the best type of assistance for the athletes to RTS in a safe and supportive manner. This literature review could also spark the interest of medical professionals to incorporate better rehabilitative practices into their work with athletes. In doing so, this will create interest in conducting new studies to see how athletes will respond to different types of rehabilitation during the process of returning after injury.

ACKOWLEDGEMENTS

The authors thank Audrey Bell- the medical illustrator in the Department of Foundational Medical Studies at OUWB- for illustrating the figure that is included in this manuscript.

REFERENCES

1. National Federation of State High School Associations. (2019). *2018-19 high school athletics participation survey. https://www.nfhs.org/media/1020412/2018-19_participation_survey.pdf*
2. National Collegiate Athletic Association. (2020). Student-athlete participation, 1981-82, 2018-19. NCAA® Sports Sponsorship and Participation Rates Report. *https://ncaaorg.s3.amazonaws.com/research/sportpart/2018-19RES_SportsSponsorshipParticipationRatesReport.pdf*
3. Woods, R. A. (2017). Spotlight on statistics: sports and exercise. U.S. Bureau of Labor Statistics. *https://www.bls.gov/spotlight/2017/sports-and-exercise/pdf/sports-and-exercise.pdf*
4. Kim, S., Bosque, J., Meehan, J. P., Jamali, A., & Marder, R. (2011). Increase in outpatient knee arthroscopy in the United States: a comparison of National Surveys of Ambulatory Surgery, 1996 and 2006. *Journal of Bone and Joint Surgery, 93* (11): 994-1000. *https://doi.org/10.2106/jbjs.i.01618*
5. Ardern, C. L., Taylor, N. F., Feller, J. A., & Webster, K. F. (2012). Return-to-sport outcomes at 2 to 7 years after anterior cruciate ligament reconstruction surgery. *American Journal of Sports Medicine, 40*:41-48. *https://doi.org/10.1177/0363546511422999*
6. Ardern, C. L., Taylor, N. F., Feller, J. A., & Webster, K. F. (2014). Fifty-five percent return to competitive sport following anterior cruciate ligament reconstruction surgery; an updated systematic review and meta-analysis including aspects of physical functioning and contextual factors. *British Journal of Sports Medicine, 48*:1543-1552. *https://doi.org/10.1136/bjsports-2013-093398*
7. Ardern, C. L. (2015). Anterior cruciate ligament reconstruction- not exactly a one-way ticket back to the preinjury level: a review of contextual factors affecting return to sport after surgery. *Sports Health: a multidisciplinary approach, 7*(3), 224-230. *https://doi.org/10.1177/1941738115578131*
8. Centers for Disease Control and Prevention. (2016). *Injury prevention and control: traumatic brain injury and concussion. https://www.cdc.gov/traumaticbraininjury/get_the_facts.html*
9. Waltzman, D., Womack, L.S., Thomas, K.E., Sarmiento, K. (2020). Trends in Emergency Department Visits for Contact Sports–Related Traumatic Brain Injuries Among Children — United States, 2001–2018. MMWR Morb Mortal Wkly Rep, 69, 870–874. *http://dx.doi.org/10.15585/mmwr.mm6927a4external icon*
10. National Collegiate Athletic Association (2014). *Mind, body and sport: understanding and supporting student-athlete mental wellness.* https://www.naspa.org/images/uploads/events/Mind_Body_and_Sport.pdf.
11. Wiebe, D. J., Comstock, R. D., & Nance, M. L. (2011). Concussion research: a public health priority. *Injury Prevention, 17*, 69-70. *https://doi.org/10.1136/ip.2010.031211*
12. Kirschen, M. P., Tsou, A., Nelson, S. B., Russell, J. A., & Larriviere, D. (2014). Ethics, law, and humanities committee, a joint committee of the American Academy of Neurology, American Neurological Association, and Child Neurology Society. Legal and ethical implications in the evaluation and management of sports-related concussions. *Neurology, 83*(4): 352-358. *https://doi.org/10.1212.wnl.0000000000000613*
13. Goldman, K. D., & Schmalz, K. J. (2004). The matrix method of literature reviews. *Health Promotion Practice*, *5*(1), 5-7.
14. Paterno, M. V., Flynn, K., Thomas, S., & Schmitt, L. C. (2018). Self-reported fear predicts functional performance and second ACL injury after ACL reconstruction and return to sport: A pilot study. *Sports Health: A Multidisciplinary Approach, 10*(3), 228–233. *https://doi.org/10.1177/1941738117745806*
15. George, S. Z., Lentz, T. A., Zeppieri, G., Lee, D., & Chmielewski, T. L. (2012). Analysis of shortened versions of the Tampa Scale for Kinesiophobia and Pain Catastrophizing Scale for patients after anterior cruciate ligament reconstruction. *Clinical Journal of Pain, 28*:73-80. *https://doi.org/10.1097/ajp.0b013e31822363f4*

16. Tripp, D. A., Stanish, W., Ebel-Lam, A., Brewer, B. W., & Birchard, J. (2011). Fear of reinjury negative affect, and catastrophizing predicting return to sport in recreational athletes with anterior cruciate ligament injuries at 1-year postsurgery. *Sport, Exercise, and Performance Psychology, 1*(5), 38–48. *https://doi.org/10.1037/2157-3905.1.s.38*

17. Shacham, S. (1983). A shortened version of the Profile of Mood States. *Journal of Personality Assessment, 47*: 305-306. *https://doi.org/10.1207/s15327752jpa4703_14*

18. Sullivan, M. J., Bishop, S., & Pivik, J. (1995). The Pain Catastrophizing Scale: development and validation. *Psychological Assessment, 7*:524-532. *https://doi.org/10.1037/1040-3590.7.4.524*

19. Vealey, R., Hayashi, S.W., Garner-Holman, M., & Giacobbi, P. (1998) Sources of sport confidence: conceptualization and instrument development. *Journal of Sport and Exercise Psychology, 20*:54-80. *https://doi.org/10.1123/jsep.20.1.54*

20. Nwachukwu, B. U., Adjei, J., Rauck, R. C., Chahla, J., Okoroha, K. R., Verma, N. N., Allen, A.A Williams, R. J. (2019). How much do psychological factors affect lack of return to play after anterior cruciate ligament reconstruction? A systematic review. *Orthopedic Journal of Sports Medicine, 7*(5), 1-7. *https://doi.org/10.1177/2325967119845313*

21. Ardern, C. L., Taylor, N. F., Feller, J. A., Whitehead, T. S., & Webster, K. E. (2013). Psychological responses matter in returning to preinjury level of sport after anterior cruciate ligament reconstruction surgery. *The American Journal of Sports Medicine, 41*(7), 1549–1558. *https://doi.org/10.1177/0363546513489284*

22. Webster, K. E., Nagelli, C. V., Hewett, T. E., & Feller, J. A. (2018). Factors associated with psychological readiness to return to sport after anterior cruciate ligament reconstruction surgery. *The American Journal of Sports Medicine, 46*(7), 1545–1550. *https://doi.org/10.1177/0363546518773757*

23. Webster, K. E. & Feller J. A. (2016). Exploring the high injury rate in younger patients undergoing anterior cruciate ligament reconstruction. *American Journal of Sports Medicine, 11*: 2827-2832. https://doi.org/10.1177/0363546516651845

24. Walker, N., Thatcher, J., & Lavallee, D. (2010). A preliminary development of the Re-Injury Anxiety Inventory (RIAI). *Physical Therapy Sport, 11*:23-29. *https://doi.org/10.1016/j.ptsp.2009.09.003*

25. Everhart, J. S., Best, T. M., & Flanigan, D. C. (2015). Psychological predictors of anterior cruciate ligament reconstruction outcomes: a systematic review. *Knee Surgery Sports Traumatology and Arthroscopy, 23*:752-762. *https://doi.org/10.1007/s00167-013-2699-1*

26. Lethem, J, Slade, D.G., Troep, J.D.G., & Bentley, G. (1983) Outline of a fear-avoidance model of exaggerated pain perception. *Behavior Research and Therapy, 21*(4): 401-408. *https://doi.org/10.1016/0005-7967(83)90009-8*

27. Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychology Review, 84*(2):191. *https://doi.org/10.1037/0033-295x.84.2.191*

28. Cohen, S. & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin, 98*(2): 310-357. *https://doi.org/10.1037/0033-2909.98.2.310*

29. Ardern, C. L., Kvist, J., & Webster, K. E. (2015). Psychological aspects of anterior cruciate ligament injuries. *Operative Techniques in Sports Medicine, 24(*1), 77-83. *https://doi.org/10.1053/j.otsm.2015.09.006*

30. Kontos, A. P., Elbin, R., Schatz, P., Covassin, T., Henry, L., Pardini, J., & Collins M.W. (2012). A revised factor structure for the post-concussion symptom scale: baseline and post-concussion factors. *The American Journal of Sports Medicine; 40*(10): 2375-2384. *https://doi.org/10.1177/0363546512455400*

31. Teel, E. F., Marshall, S. W., Shankar, V., McCrea, M., & Guskiewicz, K. M. (2017). Predicting Recovery Patterns After Sport-Related Concussion. *Journal of Athletic Training, 52*(3): 288-298. *https://doi.org/10.4085.1062-6050-52.1.12*

32. Iverson, G. (2007). Predicting slow recovery from sport-related concussion: the new simple-complex distinction. *Clinical Journal of Sports Medicine, 17*(1):31-37.

33. Slobounov, S., Slobounov, E., Sebastianelli, W., Cao, C., & Newell, K. (2007). Differential rate of recovery in athletes after first and second concussion episodes. *Neurosurgery, 61*(2): 169-173. *https://doi.org/10.1227/01.neu.0000280001.03578.ff*

34. Covassin, T., Stearne, D., & Elbin, R. (2008). Concussion history and postconcussion neurocognitive performance and symptoms in collegiate athletes. *Journal of Athletic Training, 43*(2):119-124. *https://doi.org/10.4085/1062-6050-43.2.119*

35. Covassin, T., Crutcher, B., Bleecker, A., Heiden, E. O., Dailey, A., & Yang, J. (2014). Postinjury anxiety and social support among collegiate athletes: a comparison between orthopedic injuries and concussions. *Journal of Athletic Training; 49*(4): 462-468. *https://doi.org/10.4085/1062-6059-49.2.03*

36. Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R. & Jacobs, G. A. (1983). Manual for the state trait anxiety. *Consulting Psychologists Press.*

37. Sarason, I., Sarason, B., Shearin, E., & Pierce, G. (1987). A brief measure of social support: practical and theoretical implications. *Journal of Social and Personal Relationships, 4*(4):497-510. *https://doi.org/10.1177/0265407587044007*

38. Roiger, T., Weidauer, L. & Kern, B. (2015). A longitudinal pilot study of depressive symptoms in concussed and injured/nonconcussed national collegiate athletic association division 1 student-athletes. *Journal of Athletic Training, 50*(3): 256-261. *https://doi.org/10.4085/1062-6050-49.3.83*

39. Radloff, L. S. (1997). The CES-D scale- a self-report depression scale for research in the general population. *Applied Psychological Measurement; 1*:385-401. *https://doi.org/10.1177/014662167700100306*

40. Guo, J., Yang, J., Yi, H., Singichetti, B., Starvinos, D., & Peek-Asa, C. (2018). Differences in postinjury psychological symptoms between collegiate athletes with concussions and orthopedic injuries. *Clinical Journal of Sports Medicine.* Advance online publication. *https://doi.org/10.1097/JSM.0000000000000621*
41. Kelly, K., & Erdal K. (2016). Diagnostic terminology, athlete status, and history of concussions affect return to play expectations and anticipated symptoms following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 39*(6). 587-595. *https://doi.org/10.1080/13803395.2016.1250870*
42. Cicerone, K. D., & Kalmar, K. (1995). Persistent post-concussion syndrome: The structure of subjective complaints after mild traumatic brain injury. *Journal of Head Trauma Rehabilitation, 10,* 1-17. *https://doi.org/10.1097/00001199-199506000-00002*
43. Moss-Morris, R., Weinman, J., Petrie, K., Horne, R., Cameron, L., & Buick, D. (2002). The revised illness perception questionnaire (IPQ-R). *Psychology and Health, 17*:1-16. *https://doi.org/10.1080/08870440290001494*

ABOUT STUDENT AUTHOR

Ashley Sweeney graduated from the Oakland University Honors College in the Spring of 2020. During her undergraduate career, her major was Psychology and as a requirement to graduate with honors, she completed a narrative literature review. This review was completed alongside the help of mentors Stephanie M. Swanberg and Dr. Suzan Kamel-ElSayed. This literature review is an adapted version from her Honors College thesis which is showcased in the archives at the Kresge Library at Oakland University.

She is currently furthering her education at Oakland University, pursuing her Master's in Clinical Mental Health Counseling. She is also working as a Graduate Assistant in the Office of Undergraduate Admissions on campus, where she was previously a student ambassador for three years. After graduation, she has plans to work in the business/industry setting or opening her own private practice where she grew up in the Thumb of Michigan.

PRESS SUMMARY

This narrative review aims to determine the possible psychological hindrances present when an athlete is planning on returning to sport after injury to an anterior cruciate ligament (ACL) or after sustaining a concussion. Shared psychological characteristics for returning to sport following either an ACL injury or concussion included fear, self-esteem, control, anxiety, stress, recovery, and social support. Discovering the common and unique psychological barriers which may affect the injured athletes from returning to sport can help educate athletes' families, coaches, and healthcare professionals, as well as promote discussions for the future to help athletes feel more secure in their return to their respective sport. This information can not only bring awareness to the fields of psychology, exercise science and medicine, but also further exploration into prevention of these injuries.

Print Edition ISSN 1536-4585
Online Edition ISSN 2375-8732