1. Guest Editor Comments

While clearly heading in the right direction, this paper will benefit considerably from further expositional clarity and additional content relating to the underlying themes of the Special Issue to which it is intended to contribute.  
The need for expositional clarity is raised by both reviewers, but more particularly by reviewer 1. In this vein, our view is that it is not clear the included math is adding much if anything at all – the underlying concepts seem rather intuitively obvious; is the math actually needed rather than just a slightly more considered verbal approach?

**Response**: We incorporated formula 1 and 4 in the last draft into the paragraph to avoid confusions. Per reviewer 1’s comments, we simplified the notations for current formula 1 and formula 2 in this draft.

In terms of diagrams, Figure 2 is confusing – again, is it actually needed?

**Response**: It is a good point. We decided to incorporate it into the paragraph and remove the figure.

Figure 3 is, as reviewer 2 notes, currently hard to read – and will not replicate well in grayscale.

**Response**: It is also a good point. We changed the plot to heat map.

In overall terms, more attention to the quality and suitability of the graphics will greatly enhance the approachability of the paper.  
  
The views of reviewer 2, on a need for the paper to speak more directly and substantively to an Urban Studies audience (and specifically to the special issue themes: Big Data as a tool for advancing our understanding of the urban, and to enhance urban well-being) also merit further emphasis. Attention to this matter has the potential to move the paper from a somewhat specialised contribution that will appeal to a limited cadre of transport experts, into something of much more general value and significance.

**Response**: We added many clarifications so that the paper is more suitable for the special issue theme.  
  
  
Reviewer 1  
  
The paper makes a valuable contribution in public transport research by developing intuitive quality of transfer measures based on scheduled and real-time data. The paper is well written for the most part, however there are some unclear sentences. Some minor comments are listed below:  
  
Fig 1: x-axis label further away than the legend headers, makes figure a bit hard to read

**Response**: We adjusted the graph according to the comment (see Figure 1). We indicated x and y to the corresponding label to avoid confusion; we also moved the x-axis label to the middle to keep it consistent with the y-axis label.   
  
The prime symbol used in the mathematical notation for “actual departure time” is quite difficult to notice. Can you please highlight that you use this symbol in text or alternatively use different symbols for clarity?

**Response**: We simplified the notations in the formula 1 (3.3, transfer time penalties section) and 2 (same section) in the current version. We now use uppercase T for the actual departure time and lowercase t for the scheduled departure time, which is much more obvious than prime symbols. Moreover, since all the time mentioned in the current formulas are now departure time, we removed the “departure” notations in the subscripts.  
  
Fig 2: this is not giving much additional information. It could be changed to depict all the scenarios on page 14. There should also be a scheduled arrival time for the figure to make more sense. Furthermore, the figure text could highlight that the blue line is the chosen option.

**Response**: As the reviewer 1 and editor pointed out, the figure 2 in the last draft was confusing and did not give much additional information. Therefore, we removed figure 2 and added some further explanation to the definition of each transfer type.   
  
Explanations for nonobvious abbreviations could be repeated in analysis section for convenience.

**Response**: We added several explanations in the analysis section.  
  
Are there large differences in frequency between weekdays and weekends? What are the impacts on the measures? (similar effects as on the time of day comparison?)

**Response**: We added frequency in Figure 4 and Figure 5: the points are very few (3 for weekdays and 19 for hours), so it may be trivial to add another scatter point plot. Meanwhile, it is relatively obvious to see the correlation from current line/bar plots, so we think it is better to make each plot more informative. We also added some explanation in 4.2 – hourly pattern and weekday pattern.

Weekdays – Yes, there are. COTA system have three schedules: weekdays, Saturdays, and Sundays. We can observe that frequency and the ATTP/TR measures have a positive correlation: this can be explained by the overall delays: weekdays have more delays ()

Page 23, row 50: I think it is fair to assume zero delays for demonstrative purposes, however, it would be good to also highlight that it is a highly speculative assumption that there would be no delays on a BRT line. The only conclusion that can be drawn from this simulation is that improving punctuality even on one route will reduce ATTP.

**Response**: Yes, this is a good point. We added some explanation in 4.3, paragraph 2 to highlight the assumption is hypothetical.   
  
Page 17, row 45: unclear sentence

“To investigate the spatial pattern of transfer risk, the first thing is spatial aggregation, since trip patterns (each vehicle trip; the finest level of resolution) are too specific and not representative of broader patterns.”

**Response**: Changed to: “To investigate the spatial pattern of transfer risk, the first thing is to aggregate trips based on their stops, since trip patterns (each vehicle trip; the finest level of resolution) are too specific and not representative of broader patterns.”

Page 19, row 47: unclear sentence

“However, for the APC-GTFS dataset, we observe ATTP on Sundays is second lowest compared to Fridays, which is the lowest for original GTFS dataset.”

**Response**: Changed to: “However, we observe Sundays have the lowest ATTP for the APC-GTFS dataset while Saturdays have the lowest ATTP for the original GTFS dataset.”

Reviewer 2  
The paper presents two measures of transport network performance. It is clearly written, mostly easy to follow and quite technical in nature. The editor may want to take a view to which extent the paper fits the remit of this special issue and the Urban Studies audience more generally. In its current form, the paper seems to be more suited to a transportation-focused journal.

**Response**:

Some suggestions.  
  
1. The authors could engage more thoroughly with the themes highlighted in the call of the SI. I imagine that an Urban Studies readership would be interested in a discussion of how big data can offer new understanding of urban transport systems. Of course, there are already numerous reviews on this question, but a more focused discussion in view of the SI and the particular specialism of the authors might add to the literature. This would also help readers appreciate the specific contribution of this paper.

**Response**:  
  
2. For the benefit of an Urban Studies audience, I would suggest that the authors dedicate more space to a fuller discussion of the specialist literature, and highlight the different objectives of those studies they cite. What are the pros and cons of existing measures of evaluating transfer effectiveness? In which ways are the measures proposed by authors superior to existing ones?

**Response**: TR and TTP are the first measures that provide attainable solution to quantify the real-time performance of public transit transfers.   
  
3. How does the commonly made distinction between ‘deliberate data’ and ‘byproduct data’ apply to the datasets used by the authors? I certainly agree that smartcard data are ‘byproduct data’, but datasets such as APC are deliberately collected for the purpose of passenger counts. Similarly, GTFS data are more than just byproduct; they are purposively structured, standardised and documented. The authors should clarify the ways in which their work relates to the SI’s theme of ‘Big Data in the City’, and depending on their focus, offer a fuller discussion of ‘byproduct data’ potentially extending it to issues of bias and computational cost. Alternatively, perhaps an emphasis of ‘small data’ versus ‘Big data’ may be more appropriate for this particular paper.

**Response**: We   
  
4. There is a tension between the technical measurement and passengers’ experience, which would warrant further discussion. To which extent are the components of transfer time penalties actually experienced by passengers? On high frequency services, ATP may actually not matter that much. And, if I understand correctly, transfer risk (which would be better called ‘risk of transfer loss’ or something) will increase for high frequency services. If a ‘receiving’ service runs every 5 minutes, the risk is inherently higher than if a service runs every 20 minutes, and yet the impact on the passengers’ experience is higher in the latter case. Of course, this dimension is picked up by the other measures, but I do wonder how meaningful ‘risk of transfer loss’ is without considering frequency.

**Response**: This is a good observation. I understand the question by two aspects:

1) ATP does not matter much under high frequency services.

This is true: not just for ATP, but also for all time measures. High frequency and small headway are the panacea for most public transit system. Imagine However, under

For passengers, the only factor that matters is the time penalty; in fact, as the comment pointed out, transfer risk does not matter much for them. For an individual passenger, perhaps the only thing she/he cares is the total time penalty, which is the total time loss at the receiving stop compared to the schedule.

5. Similarly, I couldn’t follow why ‘ugly, pre-emptive’ transfers would be experienced negatively. ‘Pre-emptive’ suggests that passengers pre-emptively transfer to avoid risk elsewhere; but in this particular context, they simply get on an earlier vehicle without necessarily being aware of this. Except for potential crowding, they may not be any different from ‘good’ transfers.

**Response**: Yes, it is a very good point. The reason why we called it “ugly” is not because preemptive transfers will be experienced negatively and we added some clarification in the paper about this **in 3.2 – Transfers section**.

Just like the comment says, compared to the schedule, a preemptive transfer may result in positive, or zero, or negative TTP. This is exactly the reason why we define transfer risk as the proportion of missed transfers alone, instead of missed transfer and preemptive transfers together. We acknowledged that preemptive can be as “good” as or even better than a normal transfer. However, we have to point out that preemptive transfers are different from the normal transfers. With the metaphor of “ugly”, we intended to describe its random and chaotic nature: although it may be as good, but it is not intended. Such a random merit is not sustainable.

6. P.16, last sentence, “Only those combinations with … unique routes are selected”. I couldn’t follow what unique routes means in this context. Does this mean that origin-destination pairs with multiple transfer possibilities have been excluded? More explanation would be helpful.

**Response**: We added some clarification in the text. The idea is that: for different generating stop (the stop the user gets off from the first bus) - receiving stop (the stop the user take the second bus) pairs (short for *stops pair*), if the generating stop is the same and the transfer routes pair is also the same, there is really no point for them to coexist. For example, we have *stops pairs* A->A and A->B. Stop A has bus route No. 1 and No. 2 and Stop B also has bus route No. 2, there is barely a possibility that a user will walk from A to B to catch a No.2 Bus, since she/he can always directly transfer at stop A. By doing this, we can reduce some redundancy and remove some transfers that will never be used by any users.

7. Figure 3 is hard to read. The authors may want to explore alternative types of visualisations, e.g. heat maps or contour maps.

**Response**: Will do.  
  
8. P.21 first paragraph, I wonder if statistical tests should be added to assess differences on days with rain or a football match or in the DBL scenario. By the look of the values, I don’t see how they indicate ‘considerable impact’ as the authors conclude. The authors could also focus on connections with larger differences or on those with higher passenger counts.

**Response**: Will do.  
  
9. In view again of the theme of the SI, I would expect authors to re-engage with the theme of Big Data in the conclusions. This may include a discussion of what new insights we may gain with regard to urban dwellers’ experience of their cities.

**Response**: Will do.